

**Expression analysis of a selected gene set in malignant
and non-malignant tissues derived from individuals with
colon cancer. Comparison with protein expression data.**

Dissertation

zur Erlangung des akademischen Grades
doctor rerum naturalium (Dr. rer. nat.)

vorgelegt dem Rat der Biologisch-Pharmazeutischen Fakultät
der Friedrich-Schiller-Universität Jena

von

Maria Radeva, M.Sc.

geboren am 05.02.1977 in Sofia, Bulgarien

Die vorliegende Arbeit wurde durchgeführt am:



Leibniz Institute for Age Research
Fritz Lipmann Institute (FLI)



- | | |
|---------------|---|
| 1. Gutachter: | Prof. Dr. Karl Otto Greulich
Friedrich-Schiller Universität Jena |
| 2. Gutachter: | PD. Dr. Michael Glei
Friedrich-Schiller Universität Jena |
| 3. Gutachter: | PD. Dr.rer.nat. Dr.med. Zeno Földes-Papp
University of North Texas |

Tag der öffentlichen Verteidigung: 20.07.2009

Dedicated to my family

PUBLICATIONS

Parts of this work are included in the following publications:

Radeva M, Hofmann T, Altenberg B, Mothes H, Richter KK, Pool-Zobel B, Greulich KO (2008):

The database dbEST correctly predicts gene expression in colon cancer patients. Curr Pharm Biotechnol.;9(6):510-5

Radeva M, Jahns F, Wilhelm A, Gleis M, Settmacher U, Greulich KO, Mothes H: A burst of alpha-defensin 6 expression in the adenoma stage of human colon carcinogenesis, in preparation

TABLE OF CONTENTS

TABLE OF CONTENTS	I
ACKNOWLEDGMENTS.....	IV
ABBREVIATIONS.....	VI
SUMMARY.....	VII
ZUSAMMENFASSUNG.....	VIII
1. INTRODUCTION.....	1
1.1. DIAGNOSTIC APPROACHES TO PROVE THE DETECTION OR TREATMENT OF COLORECTAL CANCER (CRC)	1
Access to tissues of different stages of colon carcinogenesis	2
1.2. COLORECTAL CANCER DEVELOPMENT	3
1.3. WNT SIGNALLING PATHWAY	5
1.4. BASIC KNOWLEDGE ABOUT THE GENES UNDER INVESTIGATION AND THE ENCODING BY THEM	
PROTEINS	7
Osteopontin (OPN).....	7
Cyclooxygenase-2 (COX-2).....	7
DEAD (Asp-Glu-Ala-Asp) box polypeptide 6 (DDX6)	8
Eukaryotic initiation factor 4E (eIF4E).....	9
Histone Acetyl Transferase 1 (HAT1).....	9
Ubiquitin Specific Protease 28 (USP28)	10
Polo-like kinase 1 (PLK1).....	10
Pyruvate kinase type M2 (PKM2).....	11
Heat Shock Proteins 90 beta (HSP90 β).....	12
Defensin alpha (DEFA 1-3 and DEFA 6).....	12
2. AIMS OF THE WORK	14
3. MATERIALS AND METHODS	15
3.1. MATERIALS	15
3.1.1. <i>Chemicals and Kits</i>	15
3.1.2. <i>Cell lines for gene expression analysis</i>	16
3.1.3. <i>Medium and reagents for cultivation of cell lines</i>	16
3.1.4. <i>Antibodies and recombinant proteins</i>	17
3.2. METHODS.....	17
3.2.1. <i>Tissue sample preparation for gene and protein expression analysis</i>	17
3.2.2. <i>Agarose gel electrophoresis</i>	19
Ethidium bromine gel electrophoresis.....	19
Formaldehyde agarose gel electrophoresis.....	19
3.2.3. <i>Gene expression analysis</i>	19
Total RNA extraction from tissue samples and cell lines.....	19
RNA integrity.....	20
Complementary DNA (cDNA) synthesis.....	20
Genomic DNA (gDNA) extraction	20
Design of primers	20
Verification of the primer specificity	21
Quantitative Real - Time PCR (qRT-PCR) conditions.....	23
PCR reaction efficiency	23
3.2.4. <i>Protein expression analysis</i>	23

3.2.4.1.	Protein extraction	23
	Preparation of lysates from tissue samples.....	23
3.2.4.2.	SDS PAGE and Western blot, quantification by ECL.....	24
3.2.4.3.	Quantification of the chemiluminescent signal by densitometry and evaluation of the Western blot data	25
3.2.5.	<i>Data mining, evaluation techniques and statistical methods</i>	25
3.2.5.1.	Data mining.....	25
3.2.5.2.	Principal Component Analysis (PCA)-the idea behind it	25
	Background of PCA. Description and terminology.....	28
	Description of the software used for PCA analysis	30
3.2.5.3.	Relative expression software tool (REST).....	31
3.2.5.4.	Statistical methods.....	32
	Non-paired, Mann-Whitney and Kruskal-Wallis test	33
	Paired, matched Wilcoxon test.....	33
	One-way ANOVA.....	33
	Linear regression and Pearson/Spearman correlation tests.....	34
4.	RESULTS	35
4.1.	QUANTITATION OF THE HOUSEKEEPING GENE (HKG) EXPRESSION IN NORMAL, NORMAL NEAR TUMOR, ADENOMA AND TUMOR COLON TISSUE ACROSS PATIENTS WITH COLORECTAL CANCER (CRC).....	35
	Housekeeping genes are stably expressed at age from 57 to 88 years	39
4.2.	NO SIGNIFICANT DIFFERENCES BETWEEN SINGLE OR MULTI-HOUSEKEEPING GENE NORMALIZATION, USING RELATIVE EXPRESSION SOFTWARE TOOL (REST)!.....	41
4.3.	INDIVIDUALITY OF THE GENE EXPRESSION AND CHANGES IN CANCER	43
4.3.1.	<i>Extremely high expression levels of DEFA 1-3 mRNAs and undetectable DEFA 6 mRNA transcription in peripheral blood mononuclear cells (PBMCs)</i>	49
4.3.2.	<i>Gene expression dependency on the age</i>	50
4.3.3.	<i>Analysis of expression levels in carcinomas according to the tumor grading</i>	51
4.3.4.	<i>Analysis of expression levels in carcinomas according to the tumor staging</i>	52
4.3.5.	<i>Gender and gene expression patterns in colorectal cancer</i>	53
4.4.	INTERPLAY OF 12 TARGET GENES IN INDIVIDUAL PATIENTS. “BAR CODE” REPRESENTATION	54
4.4.1.	<i>Discrimination of normal and cancer tissue in 18 patients with 78% accuracy and no false negatives</i>	56
4.5.	COMBINED GENE EXPRESSION ANALYSIS OF 12 GOI IN NORMAL, NORMAL NEAR TUMOR, ADENOMA AND TUMOR COLON TISSUE ACROSS 26 ANALYZED DONORS. APPLICATION OF THE RELATIVE EXPRESSION SOFTWARE TOOL (REST).....	59
4.6.	TISSUE DISCRIMINATION BY USING PRINCIPAL COMPONENT ANALYSIS (PCA).....	61
	COX-2, OPN, HSP90 β and PKM2 safely distinguished normal from tumor tissue!.....	64
4.7.	DIRECT COMPARISON OF GENE EXPRESSION DATA GENERATED BY QRT-PCR WITH THE DATA DERIVED FROM NIH’S DATA BASE DBEST.....	66
4.8.	GENE EXPRESSION IN CELL LINES	69
	<i>DEFA 1-3 and DEFA 6 expression in cell lines</i>	71
4.9.	COMBINED GENE – AND PROTEIN EXPRESSION ANALYSIS OF PAIRED NORMAL AND TUMOR TISSUE, DERIVED FROM PATIENTS WITH COLORECTAL CANCER.	74
5.	DISCUSSION	79
5.1.	β -ACTIN AND GUS AS HOUSEKEEPING GENES ARE STABLY EXPRESSED IN COLON TISSUES. NO NEED OF MULTIPLE NORMALIZATION STRATEGY!	79
5.2.	A BURST OF ALPHA DEFENSIN 6 EXPRESSION IN ADENOMA STAGE OF HUMAN COLON CARCINOGENESIS.....	80
5.3.	CORRELATION BETWEEN THE GENE EXPRESSION PATTERN AND CLINICO-PATHOLOGIC FACTORS IN COLON CANCER.....	82

5.3.1. <i>The expression of the investigated genes is not associated with the age of the patients.....</i>	82
COX-2 and ageing.....	82
PLK1 and ageing.....	83
HSP90 β and ageing.....	83
eIF4E and ageing.....	84
OPN and ageing.....	84
DEFA 6 and ageing.....	84
5.3.2. <i>The expression of the examined genes is not correlated with the grade and the stage of the cancer tissues obtained from individual patients.....</i>	85
COX-2, grade and stage of cancer.....	85
PLK1, grade and stage of cancer.....	85
HSP90 β , grade and stage of cancer.....	85
eIF4E, grade and stage of cancer.....	86
DDX6, grade and stage of cancer.....	86
OPN, grade and stage of cancer.....	86
PKM2, grade and stage of cancer.....	86
Alpha-defensins, grade and stage of cancer.....	87
5.3.3. <i>COX-2 and OPN transcriptional levels are significantly correlated with the gender of the patients under investigation.....</i>	87
5.4. <i>COX-2, OPN, HSP90β AND PKM2 ARE THE MOST PROMINENT GENES FOR DISCRIMINATION BETWEEN NORMAL AND TUMOR COLON TISSUES</i>	88
5.5. <i>GOOD PREDICTION POWER OF THE DBEST DATA BASE CONCERNING THE GENE EXPRESSION DATA OBTAINED BY QRT-PCR.....</i>	89
5.6. <i>MESSENGER RNA (mRNA) EXPRESSION DOES ONLY MODERATELY PREDICT PROTEIN EXPRESSION.....</i>	90
6. REFERENCE LIST.....	93
CURRICULUM VITAE	
SELBSTSTÄNDIGKEITSERKLÄRUNG	

ACKNOWLEDGMENTS

I am so glad to be able to say “THANK YOU!” to everyone who has been supporting me throughout the entire time I spent in Jena. I thank to all of you for loving me, accepting me and for trying to understand my difficult personality.

Firstly, I would like to express my gratitude to my supervisors - Prof Karl Otto Greulich, Prof Beatrice L. Pool Zobel (who unfortunately passed away in 2008) and Dr Michael Gleib, for their permanent interest in my work. All of them made possible this dream to come true! I would especially like to express my deep gratitude to my direct supervisor, Prof. K.O. Greulich, for giving me the opportunity to work in his department, from the beginning to the completion of this work. I thank for his continuous support, fresh ideas, “hot discussions”, honest criticism and extreme PATIENCE towards me! I learned a lot from him! He gave me great hints, brilliant tips and he was the one who introduced me to “scientific word”!

Prof Greulich, Dr. Gleib, Nelly, Chrsi, Paulius and Frau Haus were the people who stole from their spare time in order to read and correct my thesis. Their valuable comments and important advices contributed to enhance the quality of this work.

I appreciate the help of Frantziska Jahns and Kerstin Dreblow for the precise translation from English to German of my “Summary” and “Thesen”. They did a great job! Thank you ladies!

I am tremendously thankful to Paulius Grigaravicius; without his help I would not manage to perform the final formatting for such a short time.

I would like to thank all former and present members of Department “Single cell and single molecular techniques”. They all made my working-life more easy going, more pleasant and joyful- Paulius, Niki, Kerstin, Frau Haus, Shamzita, Gabi, Dr. Wollweber, Marina, Dr. Böhm, Silke, Sabine, Sven, Striebel, Alex Rapp.

I appreciate the remarkable help I received from Frau Haus and Shamzi Monejembashi (Shamzita). Frau Haus was the person who was reading every single page I was producing in English. She performed thousand of phone calls in order to help me in any matter (It is difficult to live in Germany without knowing the language). Every time I was supposed to visit any administrative office, Frau Haus was the one to join me. Many thanks to my “German mum” Shamzita, who never get tired of listening to my complains and who has always charged me with positive energy. She is one of the greatest cooks I have ever known, satisfying not only my soul, but my stomach as well. ☺

I also want to express my gratitude to Dr. Henning Mothes (Universitätsklinikum Jena) who gave me the opportunity to work with patient materials. He is one of the reasons this work could become a reality.

Special thanks to Gabi and Silke, who were constantly behind me when I was working in the cell-cultured lab!

I am grateful to Dr. Eberhard Schmitt for his unlimited help regarding the unknown for me “field of statistic”.

Thank you to Anna Kleyman, Thomas Hofmann, Raju for introducing me to the RT-PCR!

I am happy to have the opportunity to thank the people from Department of Nutritional Toxicology Institute for Nutrition, FSU, Jena – Thomas Hofmann, Julia Sauer, Nina, Frantzi, Anne, Steffi. They were so kind and friendly with me, always trying to answer my questions, being helpful and giving reasonable advices.

I am grateful that I got to know so many people from FLI, who became not only good colleagues of mine, but also nice and reliable friends. I spent a lot of time together with them- trying to relax and forget about the newly appearing professional problem.

I would like to thank Prof. Berg and his nice wife for the great time I spent with them in their garden, drinking coffee and eating cake.

Finally, I want to thank all people I love. The ones who gave me an emotional support and made my life in Jena more colorful. The ones who were with me not only in the good, but in the bad days too. They all have such a great impact in my life.

From the bottom of my heart “THANK YOU” to Nelly Traytcheva, Christina Valkova, Paulius Grigaravicius, Luchezar Karagiozov (Lucho), Nikolina Kalchishkova, and Dobromir Iliev, Boika, Franziska Jahns, Maria Traikovic, Masroor, Altaf, Hilarie, Kamran, Vertica, Ruja, Venelina Karadzova and many, many others who unfortunately due to the lack of space I can not mention here. I can not say how happy I am that faith introduced me to all of them! Their continuous support, help and company never gave me a chance give up! Whenever I had problem to be solved, one of these guys would come with a salvation! I shared with them many good moments and overcame many bad ones. Because of them, I never forget to smile and enjoy life and the people I love!

Special and deep thanks to Nelly and Christina, two amazing friends of mine, knowing me better then maybe I know myself. They both have always supported me in any occasion. Without them I would not manage to overcome many problems and bad days. Thank you girls!

I am grateful to Paulius, who has stayed with me in the lab until almost midnight, trying to discuss and solve my “scientific problems”! I respect him not only as a colleague, but also as a very good friend, who has been sharing with me a lot of time, helping me and giving me good advices. He was with me at the final stage of my thesis, right before printing. Thank you for everything Paulito!

Nikolina, Maria, Franziska and Kerstin Dreblow- four ladies with whom I shared really great moments. All of them made my life in Jena so much easier. THANK YOU!

Thanks to Lucho, a great friend, scientist and advisor not only in regard to my work, but also in regard to my private life!

Last but not least, I would like to say “THANK YOU” to all members of my big family. I want to express my deep gratefulness to my parents - Tania and Josif Radevi who dedicated their life to me and my lovely brother and sister. They have been constantly supporting me. Without them and their never ending love I would not be able to manage with this incredible challenge for me!

THANK YOU!

ABBREVIATIONS

APC	adenomatous polyposis coli
ACF	aberrant crypt foci
A	adenoma
CRC	colorectal cancer
CIN	chromosomal instability
COX-2	cyclooxygenase-2
COX-1	cyclooxygenase-1
cDNA	complementary DNA
DDX6	DEAD (Asp-Glu-Ala-Asp) box polypeptide 6
DEFA 1-3	defensin alpha 1 to 3
DEFA 6	defensin alpha 6
DMEM	dulbecco's modified Eagle's medium
eIF4E	eukaryotic translation initiation factor 4E
EDTA	ethylene diamine tetraacetic acid
EGF	epidermal growth factor
FA	formaldehyde
FAP	familial adenomatous polyposis
FCS	fetal calf serum
gDNA	genomic DNA
GOI	gene of interest
HNPCC	hereditary nonpolyposis colorectal cancer
HAT1	histone Acetyl Transferase 1
HSP90 β	heat shock proteins 90 beta
HSPs	heat shock proteins
HBSS	hank's balanced salt solution
HKG	house keeping gene
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
MCL	mantle-cell lymphoma
MIN	microsatellite instability
MMR	mismatch-repair proteins
MOPS	N-Morpholino]propanesulfonic acid
mRNA	Messenger RNA
NTC	non-template control reaction
N	normal
N nr T	normal near Tumor
OPN	osteopontin
PGE2	prostaglandin E2
PLK1	polo-like kinase 1
PKM2	pyruvate kinase type M2
PEP	phosphoenolpyruvate
PVDF membrane	polyvinylidene difluoride membrane
PCA	principal component analysis
PC	principal component
PBMCs	peripheral blood mononuclear cells
PBS	phosphate-buffered saline
qRT-PCR	quantitative real - time PCR
REST	relative expression software tool
RT	room temperature
RIN	RNA integrity number
SDS PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
T	tumor
TAE	Tris-acetate-EDTA
TE	Tris-EDTA
Tris	Tris(hydroxymethyl)-aminomethan
UTR	untranslated region
USP28	ubiquitin specific protease 28

SUMMARY

Expression of a limited set of genes (*COX-2*, *HAT1*, *USP28*, *HSP90 β* , *OPN*, *DDX6*, *eIF4E*, *DEFA 1*, *DEFA 3*, *DEFA 6*, *PKM2* and *PLK1*) was studied by quantitative Real Time-PCR (qRT-PCR) in colon tissues derived from individual patients who underwent colon surgery. The selected genes, with known tumor marker properties, code for proteins involved in cellular processes such as apoptosis, proliferation, mitosis, glycolysis, innate host defence, cellular homeostasis and translational initiation. In addition to paired-normal and tumor tissue, also pre-malignant tissues such as normal near tumor and adenoma were examined. For computation of the relative expression, β -actin was used as a reliable normalization control, because of its stable expression which was found to be independent of the age, gender or the type of the investigated colon tissue.

As known direct targets of the Wnt signaling pathway, activation mutations of which are the earliest events occurring during the cancer development, a few of the molecules, namely *OPN*, *COX-2*, *DEFA 6* and *eIF4E* were analyzed to find out whether they can be used as markers of early stages of colon cancer development and not solely, as already stated in the literature, markers of fully blown cancer. A massive expression burst of the *DEFA 6* gene in benign colon adenoma was discovered (85 fold), which immediately suggests its potential to be applied as a marker for early premalignant stages of colorectal cancer detection. Among the investigated 18 patients, the ratios in adenoma as well as in carcinoma as compared to normal tissue vary more than 1000 fold indicating the pronounced individuality of each patient's cancer. However, in 68% of the cases the overexpression in an adenoma was more than 60 fold, whereas in fully blown tumor, this elevation was much less. In almost 71% of the tumor cases the expression alteration of *DEFA 6* was less than 20 fold.

Furthermore, using an empirical approach, as well as Principal Component Analysis (PCA), 4 out of 12 tested genes, namely *OPN*, *COX-2*, *HSP90 β* and *PKM2* were determined as the most prominent ones able to distinguish non-malignant from malignant colon tissues. This finding was confirmed by a blind study where 36 tissues, obtained from 18 donors were tested. The outcome was very promising since 78% of the samples were predicted correctly and no false negatives were found.

In order to make use of the enormous information contained in NIH's data base dbEST, its predictions were directly compared with the gene expression data derived from colon tissues and generated by qRT-PCR. The analysis showed that the agreement between dbEST and qRT-PCR data is quite precise as for almost all genes the expression tendency in tumor compared to normal tissue remained the same, independent of the applied data source.

With the intention to check whether the gene expression patterns of *in vitro* models reflect the corresponding patterns of the *in vivo* counterparts, comparative analysis of gene expression data derived from adenoma and tumor tissues and cell lines (LT97, HT29 and SW480) was performed. The analysis revealed that the expression levels of *OPN*, *DEFA 1-3* and *DEFA 6* genes in cultured cells differ drastically from those observed in colon tissue samples, which should be considered to prevent any misinterpretation, when *in vitro* data is translated to *in vivo* situation.

In addition to gene expression data, quantitative Western blots were used to check as to what extent protein expression reflects the mRNA expression. *HSP90 β* , *PKM2*, *eIF4E* and *DDX6* were under investigation. Among all 12 patients analyzed, the *HSP90 β* and *PKM2* genes showed relatively low, but significant amplification in tumor tissues. On the protein side, however, no changes in the expression were found. Furthermore, when the data were evaluated in a patient-wise manner, the *HSP90 β* -mRNA and protein expression ratio between normal and tumor tissues showed to be relatively stable among the donors. In contrast, *eIF4E* disclosed higher inter-individual variability on the mRNA level, which is smoothed down on the protein level. All these observations can be a consequence of either posttranscriptional or posttranslational modifications or can be due to a smaller sensitivity of the protein assay to detect slight expression differences.

In conclusion, qRT-PCR shows that safe discrimination between normal, early and fully blown cancer is possible on the basis of approximately low number of genes and no Mega Chip is required.

ZUSAMMENFASSUNG

Die Expression ausgewählter Gene (*COX-2*, *HAT1*, *USP28*, *HSP90 β* , *OPN*, *DDX6*, *eIF4E*, *DEFA1-3*, *DEFA 6*, *PKM2* und *PLK1*) in Kolongewebe von individuellen Patienten, die sich einer Kolonresektion unterzogen haben, wurde mit Hilfe quantitativer Real Time-PCR (qRT-PCR) charakterisiert. Die ausgewählten Gene mit bekannten Tumormarker-Eigenschaften kodieren Proteine, welche in zellulären Prozessen wie Apoptose, Proliferation, Mitose, Glykolyse, angeborener Immunabwehr, zelluläre Homöostase und in der Initiation der Translation involviert sind. Neben dem gepaarten Normal- und Tumorgewebe wurden ebenso prä maligne Vorstufen, wie tumornahes Normalgewebe und Adenomgewebe, untersucht. Für die Berechnung der relativen Expression wurde β -Aktin, wegen seiner stabilen Expression, als verlässliches Referenzgen verwendet, was sich als unabhängig von Alter, Geschlecht oder Gewebetyp erwiesen hatte.

Einige der direkten Zielgene des Wnt-Signalweges (*OPN*, *COX-2*, *DEFA 6* und *eIF4E*), dessen Aktivierung durch Mutationen zu den frühesten Ereignissen in der Kolonkarzinogenese zählt, wurden analysiert, um herauszufinden, ob diese nicht nur als Marker von malignen Tumoren genutzt werden können, wie bereits in der Literatur postuliert, sondern auch als Marker für frühe Tumorstadien. Ein massiver Expressionsanstieg konnte für das *DEFA 6*-Gen in gutartigen Adenomgewebeproben detektiert werden (85fach), was das unmittelbare Potential als Marker früher, prä maligner Tumorstadien in der Entwicklung von Kolonkrebs erkennen lässt. Unter den 18 in die Untersuchungen einbezogenen Patienten, variierte die Expression im Adenom als auch im Karzinom im Vergleich zum Normalgewebe um mehr als ein 1000faches, was ein Hinweis auf die Individualität jeder Krebserkrankung ist. In 67 % der Fälle war die Expression im Adenomgewebe dennoch mehr als 60fach erhöht, wohingegen die Überexpression im malignen Tumorgewebe viel geringer war. In 71% der Proben mit Tumorgewebe war die Expression von DEFA6 weniger als 20fach verändert.

Sowohl eine empirische Erhebung als auch die Hauptkomponentenanalyse (PCA) stellten 4 der 12 untersuchten Gene (*OPN*, *COX-2*, *HSP90 β* und *PKM2*), als die Bedeutendsten dar, die eine Differenzierung von nicht-malignem und malignem Kolongewebe ermöglichen. Dieses Ergebnis konnte durch eine Blindstudie mit 36 untersuchten Geweben von 18 Spendern verifiziert werden. Die korrekte Vorhersage von 78% der Proben, ohne falsch-negative Ergebnisse verdeutlicht die Relevanz des Ergebnisses.

Um sich das umfangreiche Informationspotential der NIH Datenbank dbEST zu Nutze zu machen, wurden dessen Prognosen direkt mit den Real time-PCR Genexpressionsdaten der Kolongewebeproben verglichen. Die Analyse ergab eine recht genaue Übereinstimmung der dbEST mit den qRT-PCR Daten. Die Tendenz der Expression für nahezu alle Gene im Tumorgewebe blieb, im Vergleich zum Normalgewebe, dieselbe, unabhängig von der zugrunde gelegten Datenquelle.

Zur Überprüfung, ob das Genexpressionsmuster von *in vitro*-Modellen die entsprechenden Muster der *in vivo*-Gegenstücke widerspiegelt, wurden die Genexpressionsdaten der

Adenom- und Tumorgewebe mit denen von Zelllinien (LT97, HT29 und SW480) verglichen. Die Analyse zeigte, dass sich die Expression der Gene *OPN*, *DEFA 1-3* und *DEFA 6* erheblich von der, welche in den Kolongewebeproben festgestellt wurde, unterschied. Diese Tatsache sollte bei der Übertragung von *in vitro* Daten auf die *in vivo* Situation berücksichtigt werden, um Fehlinterpretationen zu vermeiden.

Als Ergänzung zu den Genexpressionsexperimenten wurden quantitative Western Blot Analysen durchgeführt, um zu untersuchen, in welchem Ausmaß die Proteinexpression die mRNA-Expression reflektiert. In die Untersuchungen wurden HSP90 β , PKM2, eIF4E und DDX6 einbezogen. Unter den insgesamt 12 Patienten zeigten die Gene, welche für *HSP90 β* und *PKM2* kodieren, eine relativ niedrige, aber signifikante Amplifikation in den malignen Geweben. Auf Proteinebene wurden hingegen keine Expressionsveränderungen detektiert. Bei der patienten-spezifischen Datenauswertung, zeigte sich ein relativ stabiler *HSP90 β* mRNA- und Proteinexpressionsquotient zwischen Normal- und Tumorgewebe bei den Probanden. Im Gegensatz dazu konnte eine höhere interindividuelle Variabilität auf Genexpressionsebene für *eIF4E* festgestellt werden, die auf Proteinebene allerdings nicht mehr so deutlich sichtbar war. Es ist möglich, dass die Beobachtungen entweder durch posttranskriptionelle oder posttranslationale Modifikationen hervorgerufen wurden, bzw. durch eine geringe Empfindlichkeit der Proteinanalyse-Methode geringe Expressionsunterschiede zu detektieren.

Zusammenfassend lässt sich feststellen, dass unter Verwendung von qRT-PCR eine zuverlässige Differenzierung zwischen normalem Gewebe sowie frühen und späten Tumorstadien auf Grundlage einer nahezu geringen Anzahl von Genen möglich ist und die Anwendung von Mega Chips nicht erforderlich macht.

1. INTRODUCTION

Colorectal cancer is one of the most common type of cancer with one million new cases diagnosed per year worldwide, where the developed countries account for over 65% of all cases (1).

As a prevalent disease in the ageing population, colorectal cancer develops as a multistep process from a premalignant stage via *in situ* carcinoma to invasive and metastatic cancer. Such a malignant transformation process requires years and possibly decades and it is accompanied by a number of genetic alterations such as activation of oncogenes and/or inactivation of tumor suppressor genes, see Figure 1-1, page 4. As a consequence, during the carcinogenesis many different cellular processes and signaling cascades appear to be affected, which in turn can alter the expression of various genes involved in basic processes such as apoptosis, proliferation, mitosis, glycolysis, innate host defence, cellular homeostasis, translational initiation.

1.1. Diagnostic approaches to prove the detection or treatment of colorectal cancer (CRC)

The early detection is an important factor in decreasing colon cancer deaths, therefore efficient diagnostic and therapeutic approaches are important for the success of cancer prevention. The current cancer diagnosis and classification relies on clinical and histopathological information. However, some cases bring diagnostic confusion because of incomplete clinical information or unusual histopathological features. An additional drawback of this diagnostic approach is that the histopathological information is based on criteria which are quite subjective. For that reason, the use of gene markers in addition to conventional approaches offers the opportunities for safer and improved diagnosis. The necessity to use gene expression analysis in combination with other diagnostic approaches is supported by the fact that often malignancies with the same clinical features possess different genetic alterations, which occur in different host backgrounds. Therefore the tumors can follow quite different development (2). Thus, the additional application of gene expression analysis may lead to more precise diagnosis due to the ability to identify inter-individual genetic background of tumors with the same clinical presentation and thus allowing patients to receive more adequate further treatment.

DNA microarray/chip technology is a powerful tool to investigate the complexity of cancer by measuring the mRNA expression level of thousands of genes simultaneously (2; 3). Such gene expression studies may aim to identify novel markers, reflecting the process of tumor initiation, which is desirable not only for diagnostic purposes, but for better understanding of the biological processes leading to neoplastic transformation. In addition, the assay can also be applied for testing the predictive response or resistance to treatment, especially to neoadjuvant chemotherapy (3). However, for diagnostic purposes, such a global measurement has several disadvantages. Apart from the fact that it is expensive and

therefore would be difficult to be applied in the clinical practice, this application usually possesses excessive noise generated by the often exceeding large gene set that masks the distinction of clinical outcome governed by a smaller set. In addition, due to the possible cross-hybridization between the sample and the spotted probe, the specificity of DNA mega chips can be lower compared to that when a small set of targets is analyzed in parallel. Therefore, the selection and usage of a limited set of genes with known tumor marker properties, can be a useful strategy for the development of a precise diagnostic approach. If this panel of genes safely discriminates malignant from non-malignant tissue, a small, simple, highly-specific chip can be designed for routine diagnosis, not only on gene, but on protein expression level as well.

Access to tissues of different stages of colon carcinogenesis

For performance of all diagnostic analyses, a matched pair-wise comparison of normal and cancer tissue is required. The surgical technique of resection of this tumor type results in the ability to obtain an appropriate amount of colon tissue samples for analytical studies. However, the fully blown carcinoma does not account for the expression alterations that are critical for the initiation and the development of cancer. Thus, apart from the classical combination of paired normal and tumor tissue, also tissue from intermediate, premalignant stage, namely matched normal near tumor and adenoma tissue samples should be examined. The analysis of such pre-cancerous tissues possibly would allow the discovery of markers for diagnosis of early, premalignant stages of colon cancer development, with higher sensitivity and specificity. Such an early detection is a key determinant in the survival rate of the cancer patients, since it is curable to large extent. Due to the huge variety of the human population and the high degree of heterogeneity present in the tumor, markers with high sensitivity and specificity are missing so far.

With this background knowledge, in the present work, a set of selected colon cancer markers was studied for their predictive power with respect to malignant, but also to precancerous adenoma colon tissues. Apart from the fact that all of them were detected to have enhanced expression in CRC, the chosen molecules possess different cellular functions. For a brief description of their functional properties in the cell refer to section 1.4, page 7.

Moreover, a few of the molecules investigated in the present work namely OPN, COX-2, DEFA 6, eIF4E (4-8), are known or suggested to be direct targets of the Wnt signaling pathway, activation mutations of which are the earliest events occurring during the cancer development. Therefore, it was interesting to determine (and/or confirm for some of them) whether they can be used as markers of early stages of colon cancer development and not solely, as already described, as markers for fully blown colorectal cancer. In section 1.3, page 5, short description of the Wnt pathway is given. However, the signaling cascade is not introduced in details, since it is not a goal of the present work.

1.2. Colorectal cancer development

The etiology of colorectal cancer is heterogeneous, with both environmental and hereditary factors playing roles to varying degrees. Thus, three different types of colorectal cancer with some overlapping of clinical features occur, namely sporadic, familial and hereditary cancer. Sporadic cancer is a result of the interaction between somatic mutations and environmental factors and arises stochastically and at older age. In contrast, familial cancer is clustered in families and possibly occurs due to the exposure to the some environmental risk factors or to the presence of low-penetrance mutations in susceptibility genes. On the other hand, high-penetrance germ line mutations are found in hereditary cancer (9). Thus, approximately 20% of all patients with CRC have a family history of the disease, and roughly 5-10% of the total annual burden is inherited in an autosomal-dominant fashion (10).

Hereditary cancer syndromes are divided into two categories based on the presence of polyposis, such as familial adenomatous polyposis (FAP) or hereditary nonpolyposis colorectal cancer (HNPCC). The latter one is the most common form of hereditary CRC (1). FAP is diagnosed through the presence of numerous florid colonic adenomas (the most frequent precursors of CRC carcinomas). In addition, these forms of CRC display chromosomal instability (CIN), an aneuploid karyotype, and they harbor mutations in important tumor suppressor genes and oncogenes such as APC (11; 12), K-ras and p53 (1). On the other hand, the microsatellite instability (MIN) is the hallmark of HNPCC, (13). Usually these abnormalities are repaired by mismatch-repair proteins (MMR), but their deficiency in tumors makes the repair inefficient. Mutations, caused by MIN have been found in β -catenin, proapoptotic Bax (14) and adenomatous polyposis coli (APC), (15).

However, the hereditary CRC cases are only small percentage of all CRCs and they are often detected in young age. The highest percentage of colorectal cancer cases is sporadic (non-hereditary) as they are rarely detected before the age of 40, (16).

The molecular genetics of CRC is the one of the best understood among the various human neoplasias. In Figure 1-1, page 4 the current model of carcinogenesis is illustrated (12; 17; 18).

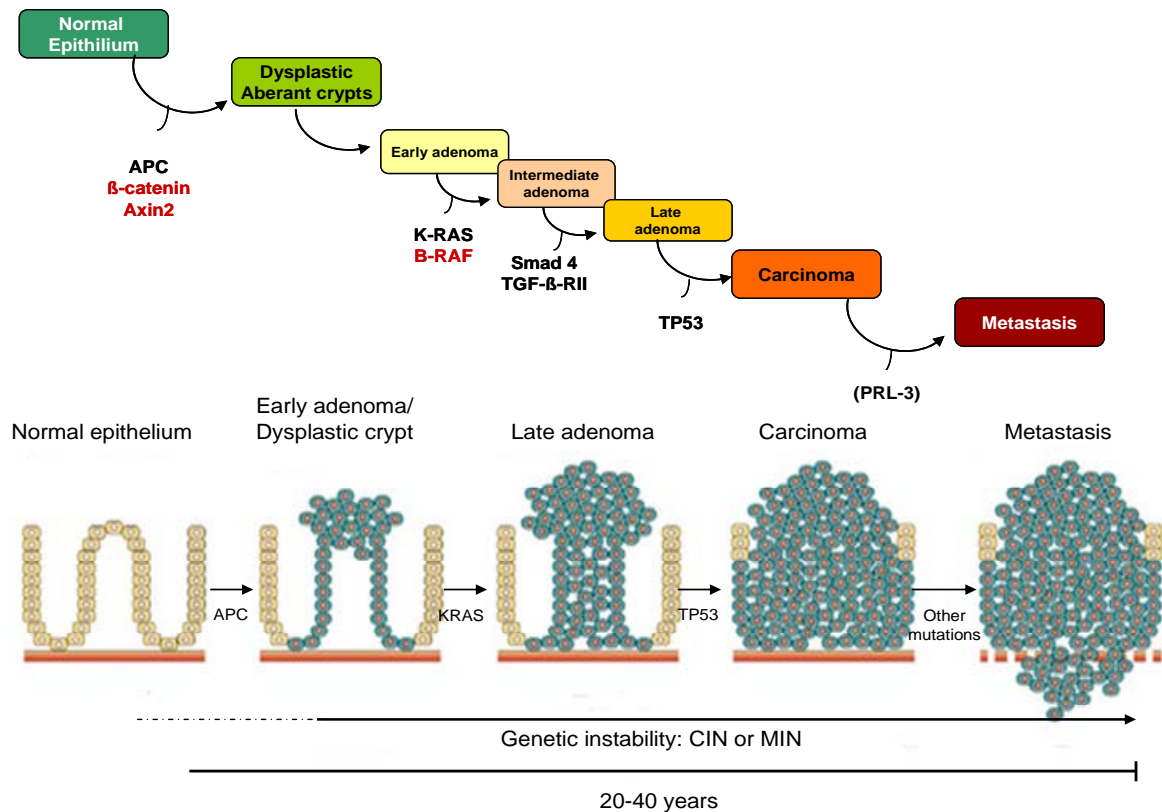


Figure 1-1 Correlation between CRC progression and the accumulation of genetic alterations according to Fearon and Vogelstein (1990). The genetic alterations frequently found in CIN tumors are depicted in black, while those depicted in red corresponds to mutations more common in MIN tumors. Such process is believed to develop over the course of 20-40 years. (modified according to (1) and (19)).

The smallest detectable lesion in the adenoma-carcinoma sequence is an aberrant crypt foci (ACF). ACFs are putative preneoplastic lesions and so far two types of ACF are distinguished (20). The most widespread type is associated with hypercellular or hyperplastic crypts that rarely turn into malignant carcinomas. The non-malignant hyperplastic forms may arise from activation mutations in K-RAS (20). The hyperplastic ACFs do not possess β -catenin mutations (21). In contrast, the second type of lesions, termed dysplastic ACF occurs often in carcinoma-associated colon mucosa and the majority of them bear mutations in APC and β -catenin molecules (21). The expansion of these dysplastic ACFs give rise to large adenomas of several centimeters. The latter one carries activating mutations in the RAS oncogene or complementary mutations in the upstream component B-RAF. Further, mutations in the TGF-beta signaling pathway (mainly in SMAD-4) confer additional malignant features to adenoma cells. Adenoma progresses into carcinoma *in situ* and in around 50% of these carcinoma cases, inactivation of TP53 is detected (22). This TP53 inactivation causes additional genetic alterations as a consequence of DNA damage and genetic instability (CIN and MIN), (1).

In addition to alterations detected in oncogenes and tumor suppressor genes, another characteristic feature occurring during the carcinogenesis is the mutation in a majority of the miss-match repair (MMR) genes (23). Apart from the entire set of genetic alterations, clear evidences revealed that the cancer is a disease driven also by epigenetic changes, such as loss or gain of DNA methylation as well as altered patterns of histone modifications (24).

1.3. Wnt signalling pathway

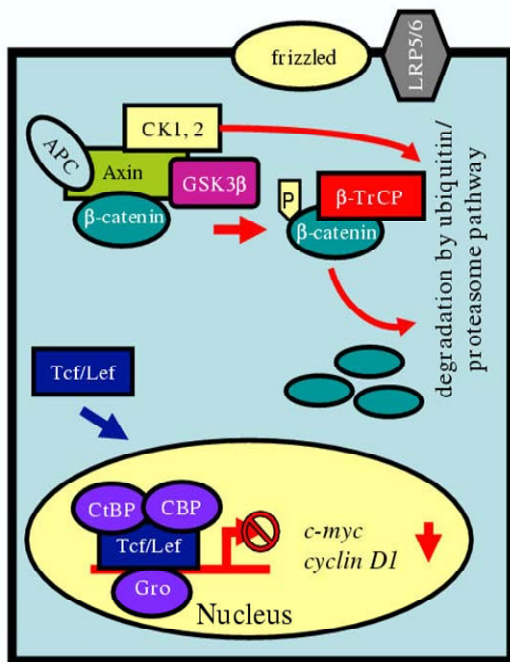
The Wnt pathway controls cell's fate during embryonic development. In addition, it also persists as a key regulator of homeostasis in adult-renewing tissues, where the mutational deregulation of the Wnt cascade was shown to be tightly associated with their malignant transformation (25).

Figure 1-2, page 6 gives a schematic illustration of how the Wnt pathway works in normal and cancer epithelial cells. For more detailed description, see the figure's legend.

Briefly: essential for this signaling pathway is the stabilization of β -catenin and its interaction with Tcf/Lef transcription factor within the nucleus. β -catenin cytosolic levels are tightly regulated as in the absence of Wnt signals, β -catenin is associated with APC, Axin and some kinases (so called APC complex), which results in its ubiquitylation and degradation by the proteosome (26; 27). Signaling by Wnt factors inhibits the APC complex. The same effect can be accomplished by mutational deactivation of APC or β -catenin, events occurring during tumorigenesis. As a consequence, β -catenin is stabilized, translocated into the nucleus, where it interacts with transcription factors driving the transcription of various target genes. List of the target genes of Wnt/ β -catenin signalling can be seen on following link:

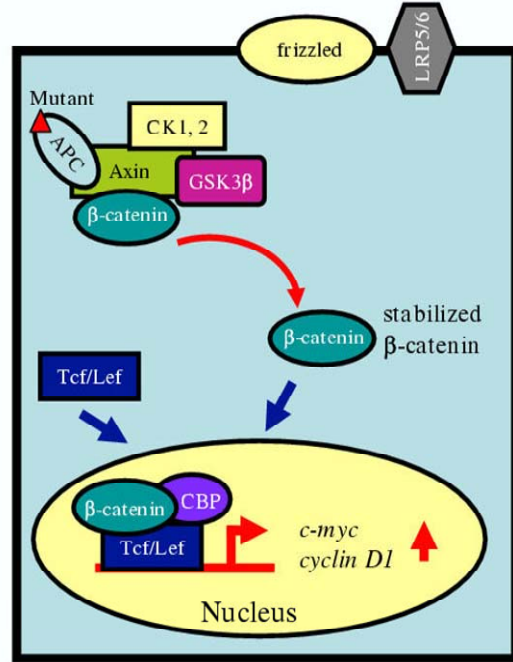
<http://www.stanford.edu/~rnusse/pathways/targets.html>

A. Normal colonic epithelial cells



Controlled cell growth

B. Colon cancer cells



Uncontrolled cell growth

Figure 1-2 The Wnt - signaling pathway. **Panel A** depicts the down-regulation of β -catenin transactivation activity in epithelial cells of healthy colon. β -catenin remains in a complex of Axin/Axil/conductin, APC, GSK3 β kinase and the casein kinase 1 or 2 (CK1 or 2). In the absence of Wnt stimulation, the kinases (GSK3 β and CK 1 or 2) become active and phosphorylate β -catenin at serine and threonine residues in the N-terminal domain. This phosphorylation is supported by APC and Axin, which acts as a scaffold protein, bringing together enzyme(s) and substrate(s). As a consequence of the phosphorylation, β -catenin binds with F-box protein β -TrCP of the Skp I-Cullin-F-box (SCF) complex of ubiquitin ligases and thus undergoes proteasomal degradation. However, Tcf-Lef transcription factor can bind to DNA, without the presence of β -catenin molecule. Nonetheless, repressors and co-repressors, namely CtBP (carboxy-terminal binding protein), CBP (CREB-binding protein) associate Tcf-Lef and suppress the transcription of cell cycle regulator molecules such as c-myc and cyclin D1. **Panel B** shows the role of the APC- and β -catenin-mutations in regard to regulation of the β -catenin level and its transactivation property in malignant colon cells. The mutant β -catenin escapes its proteosomal degradation and becomes stabilized in the cytoplasm. The same effect can be accomplished by mutational APC-deactivation. As a result, the stabilized β -catenin heterodimerizes with Tcf-Lef transcription factor and localized into the nucleus, where actively transcribes cell cycle related genes cause proliferation, survival and motility. The binding of β -catenin to TCF-Lef inhibits the binding of CtBP, CBP repressors. Figure from (28)

1.4. Basic knowledge about the genes under investigation and the encoding by them proteins

Osteopontin (OPN)

Osteopontin (OPN) is a glyco-phosphoprotein, extensively modified at both posttranscriptional and posttranslational level. As a consequence of an alternative splicing, at least three human OPN isoforms exist (29). OPN was originally isolated from bone matrix but subsequently was found in many other tissues and cell types. It is a constitutively expressed protein, but induced expression of OPN in response to various cytokines or inflammatory mediators has been detected in many cell types, including cancer cells. OPN is associated with many different cellular processes, such as cell migration and proliferation, tissue repair and angiogenesis. OPN also appears to play an important role in promoting cell survival and protecting cells from apoptosis (29-31).

Most of the cellular activities of the OPN have been assigned to the extracellular effects that occur after its secretion, where the binding of OPN to the specific receptor leads to activation of certain intracellular signaling pathways. However, an intracellular form of OPN has been previously described. It is localized in the perimembranous region of the cell and may have a specific role in the cell migration and proliferation. A recent study has shown the novel nuclear location of the intracellular OPN and its functional tandem with PLK1, also localized in the nucleus (32). Furthermore, Jain et al. have reported that OPN stimulates the activation of protein kinase C α /nuclear factor-inducing kinase/nuclear factor- κ B-dependent signaling cascades that induces Cyclooxygenase-2 (COX-2) expression (another selected target under present investigation), which in turn regulates the prostaglandin E₂ (PGE₂) production, matrix metalloproteinase-2 activation, tumor progression and angiogenesis (33). All these processes where OPN was implicated, determine its prominent role in the promotion of different types of cancers. Elevated OPN expression was detected in breast, stomach, lung, prostate, ovarian cancers and liver.(29; 31). Particularly, in numerous of studies the prominent role for OPN in the promotion of colon malignancy and metastasis has been described (34-36).

Cyclooxygenase-2 (COX-2)

There are many lines of evidence showing that PGE₂ might contribute to cancer development through various mechanisms including increased endothelial cell motility and invasion, inhibition of apoptosis, modulation of inflammation and immune response as well as promotion of angiogenesis. Elevated prostaglandin levels are also found in colon cancers and their precursor lesions, adenomatous polyps. Two forms of enzymes, Cyclooxygenase-1 (COX-1) and COX-2 are responsible for the prostaglandin's synthesis. COX-1 is constitutively expressed in most tissues and has been proposed to generate prostaglandins for normal physiological functions. The second isoform, COX-2 is widely excepted as an inducible form of an enzyme that is usually not expressed in normal tissue, but rapidly induced by tumor promoters, growth factors, cytokines, viruses and other stimuli (37; 38).

However, it was shown that in the colonic wall, COX-2 molecule is also constitutively expressed (39). So far, many studies have reported that COX-2 promotes tumor cell proliferation, survival and angiogenesis by the PGE2 mediated pathway (40; 41).

Significant increases in COX-2 expression has been found in many types of cancer as gastric, lung, pancreatic, head and neck, breast, ovaries and in other cancers (33). COX-2 expression markedly increases in 80% to almost 100% of the human colorectal carcinomas, predominantly within the neoplastic epithelial cells (38; 40; 42; 43). Enhanced expression of COX-2 in those cells is linked to alterations in the cellular adhesion and to inhibition of apoptosis (44). In addition, the constitutive COX-2 expression modulates tumor angiogenesis (45) and increases the metastatic potential of the cells (46). In contrast, COX-1 levels remain unchanged.

DEAD (Asp-Glu-Ala-Asp) box polypeptide 6 (DDX6)

DDX6 (also named as RCK or p54) is a member of the DEAD box protein/RNA helicase family. The RNA helicase activity assigned to DEAD box proteins is thought to modulate the mRNA secondary and tertiary structure (47). These proteins have been implicated in a diversity of cellular functions such as ribosome assembly, pre-mRNA splicing and translation initiation (48). It was shown that the expression of DDX6 is very poor or undetectable in brain, skeletal muscle and lung tissues, but significant amounts of this protein was found in tumors that originate from these tissues (48). Several lines of evidence presume the role of DDX6 in cell proliferation and malignant transformation by facilitating the translation of mRNA(s) for some oncogenes or growth-related genes. In this respect Hashimoto et al. have been reported that in more than 65% of colorectal tumors, DDX6 is over-expressed and that in the majority of these cases the overexpression of c-myc was also observed (49). Further, *in vitro* investigations showed that DDX6 contributes to the elevation of c-myc translation efficiency by exhibiting unwinding activity towards c-myc mRNA (50). This co-overexpression of c-myc and DDX6 is in coordination with inactivation of anti-oncogene, such as APC. The inactivation of APC gene leads to activation of c-myc transcript (51). In addition it was reported that in the early stage of colon tumorigenesis, c-myc protein elevates the expression of eukaryotic translation initiation factor (*eIF4E*) transcript, also one of the targets we are interested in. Thus, the raising of the *eIF4E* mRNA level will increase the synthesis of proteins involved in cell proliferation (8; 52). Successively, *eIF4E* increases the level of c-myc protein. Consequently, DDX6 which stabilizes and elevates the c-myc protein synthesis by unwinding the c-myc mRNA may enhance the *eIF4E* protein level. Thus, *eIF4E*, together with other translation initiation factors, may increase the protein synthesis mostly of proteins involved in cell cycle-promoting and thus leading to decrease of the cell death. As a consequence, the inability of cells to undergo apoptosis may promote cancer development.

Eukaryotic initiation factor 4E (eIF4E)

The control of mRNA translation plays a critical role in the regulation of cell growth, proliferation and differentiation. The dysregulation of this process may contribute to neoplastic transformation of the cells and consequent exhibition of malignant phenotype. eIF4E is a translation factor, the activity of which regulates the level of translation initiation. eIF4E is a cap-binding phosphoprotein, rate-limiting subunit of the eIF4F complex. eIF4E is responsible to bring the cellular mRNAs to the eIF4F complex through specific binding of cap-structure present on virtually all mature cellular mRNAs. Once engaged, the eIF4F complex can be able to scan the mRNA in 5'-3' direction from the cap, unwinding any secondary structure within the 5'untranslated region (UTR) to uncover the translation initiation codon and to facilitate ribosome loading on the mRNA (53).

In normal living cells, the accessibility of eIF4E is limited and as a consequence the translation is restricted. Under these conditions, where the molar concentration of mRNA exceeds that of eIF4E, the mature mRNAs have to compete for eIF4E in order to win access to ribosome machinery. There are two types of mRNA, namely highly and weakly competitive mRNA. The highly competitive mRNAs have a short, unstructured 5'UTR and therefore their scanning proceeds easily from 5' cap through the short leader sequence in order to reveal the initiation codon and thus to facilitate the translation process. Such a well translated mRNAs are the majority of the cellular mRNAs (e.g. β -actin). In contrast, the less competitive mRNAs have highly-structured 5'UTR, which make them poorly translated in normal cells. These mRNAs encode proteins that stimulate processes; well know to be deregulated in cancer; such as growth and angiogenesis (e.g., fibroblast growth factor 2 (FGF-2), vascular endothelial growth factor (VEGF), cyclin D1 and c -myc) as well as cell survival (e.g Bcl-2) and invasion (e.g. MMPs). However, the translation efficiency of the poorly translated mRNAs can grow only under condition where eIF4E activity is elevated (53; 54).

Increased expression of eIF4E has been found in broad spectrum of human cancers and cancer cell lines, including bladder, breast, cervical, head and neck, and prostate (53-55). Additionally, numerous reports described that the initiation factor eIF4E is strongly involved in colon tumorigenesis (8; 54). Elevated eIF4E expression and activity may contribute to increase in both, general protein synthesis and preferential synthesis of specific growth-promoting proteins. Therefore, it is hypothesised its major potential role in tumor progression. Due to this fact, it might be a useful intermediate biomarker for use in chemoprevention intervention studies in patients with colorectal polyps.

Histone Acetyl Transferase 1 (HAT1)

In the intact cell, DNA is closely associated with histones and other nuclear proteins to form chromatin. By posttranscriptional modification of the nucleosomal histones, the regulation of the gene expression in the eukaryotic organism can be performed. One such covalent modification is the acetylation of the lysine residues in the amino terminal tail domain of highly conserved histones. This alteration leads to conversion of the chromosomes's regions into transcriptionally active chromatin since the accessibility of transcription factors to the

DNA template has been increased. The process is catalyzed by histone acetyltransferases, which are separated in two types, type A and B, depending on their subcellular localization and function.

Type A HATs are located in the nucleus. Many of them play a role as transcription co factors, to be able to regulate the gene expression.

HAT1 is type B enzyme. Type B HATs are thought to be located in the cytoplasm and acetylate newly synthesized histones, but not nucleosomal histones, prior to chromatin assembly (56). The acetylation of the nascent histones would be expected to promote transcription by facilitating the binding of transcription factors to nucleosomal DNA (57). Interestingly, there are few reports revealing the nuclear localization of HAT1 in S-phase cells too (56; 58). HAT1 may play role in the telomeric silencing (59).

Substantial increase in the expression of HAT1 was detected in primary and metastatic colon tumors, in liver tumors as well as in a variety of other cancers (60-62).

Ubiquitin Specific Protease 28 (USP28)

MYC as a key regulator of cell growth, proliferation and apoptosis is implicated in the genesis of many human tumors. Increase in its gene- and protein-expression has been observed in a majority of the colorectal adenomas compared to their adjacent, normal mucosa (63). Stabilization of MYC proto-oncogene can be achieved by USP28, which is antagonizing the activity of SCF-FBW7 complex. The latter is responsible for MYC degradation through the proteasome pathway (64).

Consistent with the necessity for high level of MYC for proliferation of some human cancer cells, Popov et al. have reported that the expression of USP28 was elevated in colon and breast carcinomas to enhance the stability of MYC protein. In contrast, they have shown that the reduction in the USP28 expression can strongly inhibit the growth of the cancer cells (64).

Polo-like kinase 1 (PLK1)

PLK1 is a member of the serine/threonine protein kinase family, cdc5/polo subfamily. It has been shown that constitutively expressed PLK1 is a mitotic regulator, responsible for phosphorylation of Cdc25C and cyclin B1 molecules, which leads to their activation and nuclear translocation. The nuclear accumulation of Cdc25C may promote the activation of cdc2/cyclin B1 complex in the nucleus by counteracting the inhibitory activity of Wee1, which is constitutively nuclear. This event is thought to be important for driving the cells into mitosis (65; 66). PLK1 is also involved in centrosome maturation and assembly, spindle function, mitotic exit and cytokinesis and therefore its inactivation may contribute to mitotic arrest, induction of the pro-apoptotic pathway and suppression of tumor growth in response to stress (67). In contrast, PLK1 expression positively correlates with the expression of markers for cell proliferation, such as Ki-67 (68). PLK1 overexpression may lead to elevated proliferation and cellular transformation. Smith et al. have shown that the overexpression of PLK1 in murine fibroblasts increased the number of the cells with fragmented or multiple nuclei. When these cells, with typical phenotype of transformed

cells in culture were injected into mice, tumor could be caused (69). Moreover, recent study has shown the interaction of OPN with PLK1, which is limited in the nucleus, presumably during mitosis. As PLK1, elevated expression of OPN also can promote the multinucleation in different cell types. This common biological role, together with the founded physical interaction between OPN and PLK1 supports the speculation that these two targets function together by taking a part in the process of cell duplication (32).

Furthermore, enhanced expression of PLK1 has been reported in various types of cancer as lung, skin, breast, prostate, stomach, head and neck, brain, ovary as well as colon cancer and often it is correlated with poor patient prognosis (68; 70). Many lines of evidence defined the PLK1 expression as a suitable marker of metastasis (71). It was also reported that by phosphorylation of p53 tumor suppressor protein, PLK1 inhibits the pro-apoptotic functions of p53 (72).

Pyruvate kinase type M2 (PKM2)

One common feature during the development of cancer is the alteration in the expression of enzymes involved in the glycolytic pathway. The protein encoded by *PKM2* gene is a pyruvate kinase, a glycolytic enzyme which plays a key role in this pathway. The pyruvate kinase catalyzes the dephosphorylation of phosphoenolpyruvate (PEP) to pyruvate and it is responsible for the ATP-production. Several different isoenzymes from this kinase are identified (L, R; M1; M2, tumor M2). They are tissue-specifically expressed in different organisms. L type is the major isoenzyme in the liver, R is found in red cells (erythrocytes), M1 is the main form in muscle, heart and brain, and M2 is predominant in lung tissue as well as in cells with increased nucleic acid synthesis, as for example proliferating cells like those from the fetal tissues, adult stem cells and tumor cells (73; 74). All these isoforms occur as a homotetramers in their active state with high affinity to their substrate PEP. During the tumorigenesis the tissue specific isoenzyme, such as PKM1 in brain or PKL in the liver, generally disappears and the level of PKM2 strongly increases. As a consequence PKM2 releases into the blood and in the case of gastrointestinal tumors also into the stool of tumor patients possibly because of tumor necrosis and cell turnover (74; 75). Interestingly, in contrast to the normal proliferating cells, where PKM2 is mainly found as tetramer (74), the isoform of PKM2 found in tumors is in dimeric state, and it is termed tumor PKM2. This isoform has lower PEP affinity and therefore all glycolytic intermediates above pyruvate kinase are accumulated and channelled into synthetic processes as for example nucleic acid synthesis, amino- acid and phospholipids synthesis (76).

The increased expression of PKM2 can be caused by mutation in the ras gene (in more than 40 to 60 % of the colon cancer patients ras has been shown to be mutated) (76) and up-regulation of transcription factors like HIF1, SP1 and SP3 (74). Furthermore, the dissociation of the tetrameric form to dimeric form in cancer cells mainly is caused by the interaction of PKM2 with oncoproteins as pp60 v-src kinase and HPV-16 E7 (77).

It was reported that the level of tumor PKM2 correlates with the staging and the ability of the tumor cells to metastasise. Immunohistological analysis has shown the heterogeneous distribution of tumor PKM2 in various primary tumors while their metastases are

characterized with homogeneously large amount of tumor PKM2 (78; 79). In agreement with this data, elevated expression of tumor PKM2 has been found in EDTA-plasma samples of patients with pancreatic, lung, renal, breast and gastrointestinal carcinoma as well as in stool samples of patients with colorectal and stomach cancer. Moreover, the PKM2 expression is not organ specific and reflects the metabolic activity of the tumor (74).

Heat Shock Proteins 90 beta (HSP90 β)

HSP90 β is a member of one of the most evolutionary conserved classes of molecules called Heat Shock Proteins (HSPs) or Stress Proteins. They play a fundamental role in the maintenance of cellular homeostasis. Under normal physiological conditions, HSPs act as "molecular chaperones". As such, they are critical to maintaining the normal protein-folding environment and therefore are able to modulate the protein activity. In addition, these chaperones can affect assemble/disassemble of the protein complexes, the protein translocation or degradation through the proteosome pathway and thus regulate the apoptosis (80). Their elevated expression increase the ability of the cell to survive in tissues harmed by different types of stress such as heat, heavy metals, hypoxia, or acidosis. These conditions are commonly occurring during tumorigenesis. Therefore their altered usage during oncogenesis is critical to the development of human cancers by reflecting the ability of malignant cells to maintain homeostasis in a hostile environment, but also allows the tumor cells to tolerate alterations from within, including mutation of critical signaling molecules that would otherwise be lethal (80; 81). HSP90 can promotes the initiation of angiogenesis (82). Additionally, HSP90 inhibits apoptosis (83).

HSP90 family members are HSP90 α and HSP90 β , but the functional differences between these two isoforms are still poorly understood as in most expression studies, differentiation between α and β isoforms of HSP90 was not performed. However, several reports described that the abundant HSP90 protein overexpressed in Mantle-cell Lymphoma (MCL), leukemias, Hodgkin's disease, as well as in breast and lung cancers (84). In addition, Hui Cen et al. showed that the expression of HSP90 was increased in colon cancer tissue in comparison with normal colon tissue by using PathwayFinder GEArray (85).

Defensin alpha (DEFA 1-3 and DEFA 6)

Human defensins are a family of naturally occurring cationic short peptides of 29 to 42 amino acids in length, which have shown, in addition to their well known diverse functional activities in innate antimicrobial immunity (86-88), elevated expression in various tumor types, including colon cancer. In contrast to DEFA 1-3 which has been shown to be an unspecific colon cancer marker, DEFA 6 is more tissue specific due to its high expression in colon cancer as compared to other tumors (89-93).

In humans, two subfamilies, namely alpha- and beta-defensins exist. The known six alpha-defensins include the neutrophil defensin alpha 1 to 4 (DEFA 1-4) and the enteric DEFA 5 and 6. DEFA 1-4 are major components of the dense azurophilic granules of neutrophils, while DEFA 5 and 6 are primarily expressed in the lysozyme-rich granules of the Paneth cells of the small intestine, but also found in intermediate cells (94; 95).

In contrast to DEFA 1 and 3 peptides which differ only by the first amino acid and are encoded by different genes, so far no gene for DEFA 2 was discovered. Therefore, it is thought that DEFA 2 is a proteolytic product of one or both of DEFA 1 and DEFA 3 peptides (96).

Independent of the origin of the defensin molecules, it is essential to reveal the function they have in the tumor microenvironment, *in vivo*. Depending on the amount of defensin in the tumor tissue, *in vitro* analyses illustrated that different modulations of the cancer growth can be promoted. At physiological concentration, defensins can assist the cancer survival via exhibiting mitogenic effects (97). It was also shown that defensins can endogenously bind to HLA-DR, a MHC II molecule on hematopoietic progenitor cells and renal carcinoma cells, and therefore can restrict the recognition of the cancer cells by T-cells (91; 98). In addition, defensin may negatively regulate NK cells and CD4+ T cells, which can help the cancer cells to avoid the local immune recognition (99). This activity, together with the mitogenic effect can favor malignant growth and progression (97).

In contrast, at high concentration, addition of defensins can exert cytotoxic effects (91; 97). In this respect, it was already reported that tumor cells located at regions with extended defensin patches tended to display morphological signs of necrosis (93). Müller et al. also revealed that the strong labeling of DEFA 1-3 in some parts of the tumor seemed to be correlated with large areas of cellular necrosis (97).

Recent publication showed that DEFA 1 molecule was expressed by the tumor cell and as a consequence this intracellular localization can directly results in tumor cells-apoptosis with considerable inhibition of tumor growth *in vivo*. In addition, the potential antitumor effect of DEFA 1 was supported by the observed inhibition of tumor angiogenesis (100). In contrast to these observations, another report revealed that apart from the verified alpha-defensin expression in bladder cancer cells exposure to this molecule increased the motility/invasiveness and proliferation of tumor cells *in vitro* (101).

Taking into consideration all above described observations it can be suggested that alpha defensin molecules potentially can modulate colorectal cancer progression. However, the question whether they pro- or demote this process remains unclear.

2. AIMS OF THE WORK

To find out:

- the most informative combination of selected genes able to distinguish normal from tumor colon tissues derived from patients having colon surgery.
- whether some of the molecules, known as targets of the Wnt pathway, mutational activation of which is the earliest event occurring during colon cancer development, can be used as a markers for early detection of colorectal cancer.
- whether the gene expression patterns of *in vitro* models, such as LT97, HT29 and SW480 colon cancer cell lines, reflect the corresponding patterns of the *in vivo* counterparts, namely different types of colon tissues.
- whether the wealth of information contained in NIH's data base dbEST reflects the data derived from human individuals by quantitative Real-Time PCR.
- to what extent the protein expression pattern of the analyzed molecules reflects the corresponding gene expression pattern.

3. MATERIALS AND METHODS

3.1. Materials

3.1.1. Chemicals and Kits

Acetone	Roth, Karlsruhe, GER
Acryl-Bisacrylamid (29:1)	Roth, Karlsruhe, GER
Agarose Broad Range	Roth, Karlsruhe, GER
Ammonium persulphate	Sigma-Aldrich, Munich, GER
B-glycerolphosphate	Sigma-Aldrich, Munich, GER
Bovine Serum Albumine (Fraction V)	Sigma-Aldrich, Munich, GER
Bromphenol Blue	Sigma-Aldrich, Munich, GER
Butanol	Roth, Karlsruhe, GER
DEPC	Roth, Karlsruhe, GER
DMEM	Promocell GmbH, GER
DNase I, Amplification Grade	Invitrogen GmbH, Karlsruhe, GER
dNTRs mix	Promega, GER
DTT	Sigma-Aldrich, Munich, GER
ECL Pierce Western Blotting Substrate	Perbio Science, Bonn, GER
ECL Plus Amersham Western Blotting Detection System	GE Healthcare, Buckinghamshire, GBR
EDTA	Roth, Karlsruhe, GER
EGF (epidermal growth factor)	Calbiochem-Novabiochem
Ethanol	Roth, Karlsruhe, GER
Ethidium Bromide	SERVA, Heidelberg, GER
Fetal calf serum	PAA Laboratories, Pasching, AUT
Formaldehyde 37%	Sigma-Aldrich, Munich, GER
Formamide	Sigma-Aldrich, Munich, GER
Gentamicyn	Invitrogen GmbH, Karlsruhe, GER
Glycerol/ glycerine	Roth, Karlsruhe, GER
Go Tag DNA Polymerase	Promega, GER
Hybond-ECL Nitrocellulose Membrane	GE Healthcare, Buckinghamshire, GBR
Hyperfilm ECL Amersham	GE Healthcare, Buckinghamshire, GBR
Hydrocortisone	Sigma-Aldrich, Munich, GER
Isopropanol	Roth, Karlsruhe, GER
iQ™ SYBR® Green Supermix	Bio-Rad, GER
Insulin	Sigma-Aldrich, Munich, GER
KCl	Sigma-Aldrich, Munich, GER
KH ₂ PO ₄	Merck, Darmstadt, GER
L-15 Medium (Leibovitz)	Sigma-Aldrich, Munich, GER
Methanol	Roth, Karlsruhe, GER
2-mercaptoethanol	Merck, Darmstadt, GER
MCDB 302	Biochrom AG, GER
MgCl ₂	Sigma-Aldrich, Munich, GER
MOPS	Roth, Karlsruhe, GER
Na Acetate	Sigma-Aldrich, Munich, GER
NaCl	Roth, Karlsruhe, GER
NaF	Roth, Karlsruhe, GER
Na ₂ HPO ₄	Merck, Darmstadt, GER
NaH ₂ PO ₄	Roth, Karlsruhe, GER
NaOH	Roth, Karlsruhe, GER
Na ₃ VO ₄	Sigma-Aldrich, Munich, GER
NEA-non essential amino acids	Biochrome AG, GER
Nonfat dried milk powder	AppliChem, Darmstadt, GER
NP40	Sigma-Aldrich, Munich, GER
Oligo(dT)12-18 primers	Invitrogen GmbH, Karlsruhe, GER
PageRuler Prestained Protein Ladder Plus	Fermentas, St. Leon-Rot, GER
Penicillin/Streptomycin Solution	Sigma-Aldrich, Munich, GER
PCR mycoplasma test kit	AppliChem GmbH, GER
PMSF	Sigma-Aldrich, Munich, GER
PonceauS Concentrate	Roth, Karlsruhe, GER
Protease Inhibitor Cocktail	Roche, Mannheim, GER
Pefabloc	Roth, Karlsruhe, GER
QIAquick Gel Extraction Kit	QIAGEN, Hilden, GER
RNase H	New England Biolabs

RNAlater	Qiagen, Hilden, GER
Rneasy® Mini Kit	QIAGEN, Hilden, GER
Rotiphorese blauR brilliant blue concentrate	Roth, Karlsruhe, GER
Roti®-Quant (concentrate)	Roth, Karlsruhe, GER
Rotiphorese Gel 40 (19:1)	Roth, Karlsruhe, GER
RPMI 1640 Medium	Biochrom AG, Berlin, GER
SDS	SERVA, Heidelberg, GER
Smart Ladder (Dann Standard)	Eurogentec, Köln, GER
Smart Ladder Small Fragment	Eurogentec, Köln, GER
Sodium Azide	Sigma-Aldrich, Munich, GER
Sodium selenite	Sigma-Aldrich, Munich, GER
SuperScript II Reverse Transcriptase	Invitrogen GmbH, Karlsruhe, GER
TEMED	Roth, Karlsruhe, GER
Tris Base	Sigma-Aldrich, Munich, GER
Transferrin (Esther)	Invitrogen GmbH, Karlsruhe, GER
Tris HCl	Roth, Karlsruhe, GER
Trypsin/ EDTA Solution	Biochrom AG, Berlin, GER
Tween 20	Roth, Karlsruhe, GER
Trijodo-L-Thyronine	ICN
Xylene cyanol	Sigma-Aldrich, Munich, GER

3.1.2. Cell lines for gene expression analysis

Six cell lines were examined in the present study. Table 3-1 provides general information about the all cell lines. Five of them (HT29, Caco 2, HCT116, SW480 and SW620) were established from primary colonic adenocarcinomas. These cell lines differ greatly in their degree of differentiation, their proliferation rate, and their metastatic potential. One cell line (LT97) was isolated from colonic microadenomas.

Table 3-1 Description of tumor cell lines

Cell line	Description
LT97	LT97 are early adenoma cells isolated from microadenomas of a patient suffering from hereditary familial polyposis. They possess genetic alteration characteristics of early adenoma <i>in vivo</i> . LT97 cells have lost both alleles of the APC tumor suppressor gene (102).
HT29	HT29 is a moderately well differentiated human colon cancer cell line which was established from a grade II tumor. When injected into nude mice, it produces well-differentiated tumors consistent with grade I tumors (103). HT29 contain mutant APC protein (104).
Caco 2	Forms moderately well differentiated adenocarcinomas consistent with colonic primary grade II, in nude mice (105). Caco 2 contain mutant APC protein (104).
HCT116	HCT116 is reported to have the microsatellite instable (MSI) phenotype and it is highly tumorigenic (106). HCT116 contain wild type APC protein (104).
SW480	SW480 is a poorly differentiated cell line which was established from a grade III/IV primary tumor. A lymph node metastasis arising from this tumor was isolated and used to generate the SW620 cell line (103). SW480 contain mutant APC protein (104).
SW620	See SW480. They contain mutant APC protein (104).

3.1.3. Medium and reagents for cultivation of cell lines

HT29, SW480 and SH-SY5Y cells were routinely grown in RPMI 1640 Medium, with stable glutamine, supplemented with 10% Fetal Calf Serum (FCS), 10mM β -mercaptoethanol and 1% penicillin/streptomycin solution.

LT97 cell line was maintained in a culture medium (MCDB 302), containing 20% of L15 Leibovitz medium (supplemented with L-Glutamin) ; 2% FCS, 0.2nM triiodo-L-thyronine; 2 µg/ml transferrin; 1 µg/ml hydrocortisone; 10 µg/ml insulin; 5nM sodium selenite and 30 ng/ml EGF (epidermal growth factor) and 1% penicillin/streptomycin solution.

Caco 2, HCT116 and SW620 were cultured in Dulbecco's modified Eagle's medium (DMEM) with stable glutamine and 1% penicillin/streptomycin solution. For Caco 2 and HCT116, DMEM was supplemented with 20% FCS and 0.5% β-mercaptoethanol. In the case of SW620, 1% v/v non-essential amino acid solution and 20% v/v of FCS were added to the medium.

The cells were usually grown in 25 cm² or 75 cm² tissue culture flasks to 75-85% confluence. The cultivation was carried out in a humidified incubator (5% CO₂; 95% humidity, at 37°C). Cells were washed twice with PBS, trypsinised (Trypsin/EDTA) and after centrifugation at 80 g for 5 min the cell pellets were collected and resuspended in appropriate lysis buffer either for further RNA or protein extraction.

3.1.4. Antibodies and recombinant proteins

The antibodies and recombinant proteins, used in this work are listed in Table 3-2 and Table 3-3, respectively.

Table 3-2 List of antibodies

Antibody name	Ig Type	Company	Used dilution
Anti – COX-2	Goat monoclonal,	R&D systems	1.5 µg/m
Anti – eIF4E	Mouse monoclonal	BD Transduction Lab	1: 500
Anti – PKM2	Rabbit policlonal	ABGENT	1: 500
Anti – DDX6	Rabbit policlonal	Bethyl Laboratories	1: 10 000
Anti – β Actin	Mouse monoclonal	SIGMA	1:10 000
Anti – HSP90β	Mouse monoclonal	Zymed Laboratories	1.5 µg/ml
Anti – Rat	Goat peroxidase conjugate	Santa Cruz	1: 500
Anti – Rabbit	Goat peroxidase conjugate	Dako Cytomation	1: 10 000
Anti – Mouse	Goat peroxidase conjugate	SIGMA	1:10 000 or 20 000
Anti – Goat	Donkey peroxidase conjugate	Santa Cruz	1:5 000 or 10 000

Table 3-3 List of recombinant proteins

Protein name	Molecular Weight	Company
Recombinant human eIF4E	28 kDa	GloboZymes
Recombinant human COX-2	70-74 kDa	Bizol
Recombinant human COX-2	70-74 kDa	Oxford Biomedical Research
Recombinant human HSP90	90 kDa	Alexis

3.2. Methods

3.2.1. Tissue sample preparation for gene and protein expression analysis

All patients have given their informed consent and the study was institutionally approved by the ethics committee of the Friedrich Schiller University of Jena. Altogether, tissue samples

from 36 patients were included in the study as 6 of them are examined by Frantziska Jahns (Department of Nutritional Toxicology, Institute for Nutrition, Friedrich-Schiller-University Jena), in respect to the early phase of her PhD work.

The donors of the colon tissue were admitted to the University Hospital of Jena, Germany to undergo surgery for removal of colon tumors or colon polyps. The normal colon samples from each patient with colon cancer were taken at a distance of 20 to 50 cm from the tumor site. Confirmation of the tumor stage of the patients was provided by pathological examination after the surgery (Table 3-4). In addition to the adenoma samples, macroscopically normal tissues, removed at location 1 to 5 mm near by the tumor were collected. All adenoma samples were benign.

The tissue samples were stored in Hank's balanced salt solution (HBSS) and transported to the laboratory on ice within 1 h after the surgery. The human colon epithelium was separated from the tissue by perfusion-supported mechanical disaggregation (107). The epithelial stripes were either immediately submerged in RNeasy lysis buffer, stored at -80°C until further use for RNA extraction or were placed in 1.5 ml plastic tubes, put in liquid nitrogen and stored at -80°C for subsequent protein extraction.

Total RNA from leukocytes (peripheral blood mononuclear cells (PBMCs)), derived from eight healthy donors was kindly provided by Thomas Hofmann (Department of Nutritional Toxicology, Institute for Nutrition, Friedrich-Schiller-University Jena).

Buffers:

Hank's balanced salt solution (HBSS) – 0.8 g/L NaCl, 0.4 g/L KCl, 0.06 g/L Na₂HPO₄ x 2 H₂O, 0.06 g/L K₂HPO₄, 1 g/L glucose, 0.35 g/L NaHCO₃ and 4.8 g/L HEPES, pH 7.2)

Table 3-4 Clinicopathological characteristics of 36 patients. Four different types of tissues were examined for expression analysis of 12 genes implicated in various cellular processes as well as 2 reference genes.

Patient and tumor characteristics	Number of cases	
Type of analyzed tissue		* One of the analyzed patients had two types of benign adenoma
Normal	36	
Normal near Tumor	18	
Adenoma	19*	
Tumor	33**	
Gender		** One of the donors did not possess cancer colon tissue
Male	24	
Female	12	
Age (years)	Mean age 71 ± 9.4	
Tumor grade		*** Information on tumor grade was not available
GX	1***	
G1	3	
G2	14	
G3	14	
G4	1	
Tumor stage (according to TNM system)	33 patients in total	
UICC I	8	
UICC II	8	
UICC III	11	
UICC IV	6	

3.2.2. Agarose gel electrophoresis

Ethidium bromide gel electrophoresis

In order to confirm the specificity of the designed primer pairs, size verification of the amplified PCR products was performed. The PCR mixtures were combined with DNA loading buffer and separated using 1.2-2% ethidium bromide stained agarose gels. The agarose was dissolved in TAE buffer. Gels were run at 100 V and the products were visualized under UV light and documented with Argus X1 software (Biostep GmbH, Germany). The length of the DNA fragments was determined by comparing their electrophoretic mobility with ready-to-use DNA marker sample of known lengths (range of the size 100-1000 bp).

Buffers:

DNA loading buffer – 50 mM Tris-HCl, pH 7.6, 30% Glycerin, Xylene cyanol, Bromophenol blue

TAE buffer – 40 mM Tris, 20 mM acetic acid, 1 mM EDTA, pH 8.0

Formaldehyde agarose gel electrophoresis

To determine the integrity of the isolated total RNA, sharp bands of the respective ribosomal RNAs should appear after running the samples on formaldehyde agarose gel (FA gel). Prior to be loaded on a gel, the samples were mixed with 1x RNA loading buffer and boiled at 65°C for 5 min. The apparent ratio of 28S rRNA to 18S rRNA should be approximately 2:1. If the observed ribosomal bands are smear, it is likely that the samples are degraded before or during the RNA purification.

Buffers:

10x FA gel buffer – 200 mM MOPS, 50 mM Sodium Acetate, 10 mM EDTA, pH 7 (with NaOH)

1x FA gel running buffer – 100 ml, 10x FA gel buffer, 20 ml, 37% formaldehyde, 880 ml RNase-free water

1x RNA loading buffer – 16 µl saturated bromophenol blue solution; 80 µl, 500 mM EDTA, pH 8; 720 µl 37% formaldehyde; 2% glycerol; 3.084 µl formamide, 4 ml 10x FA gel buffer; RNase-free water to 10 ml

3.2.3. Gene expression analysis

Total RNA extraction from tissue samples and cell lines

Total cellular RNA was isolated either from epithelium colon stripes or from cell pellets derived from different cell lines using RNeasy Mini Kit where genomic DNA contaminations were effectively removed by using genomic DNA Eliminator spin column. Proper volume of RLT Plus lysis buffer was added to the fresh prepared cell pellets and/or tissue samples. The cells were homogenized using a syringe and needle, while for the deep frozen colon stripes Polytron PT 2100 homogenizer (Kinematica AG, Littau-Luzern, Switzerland) was used. Next steps of the RNA isolation were performed according to the manufacture's instructions. The RNA yield was determined by measuring the absorbance at 260 nm (A_{260}) in a NanoDrop®ND-1000 Spectrophotometer (NanoDrop Technologies,

Montchanin, DE, USA). An absorbance of 1 unit at 260 nm corresponds to 44 µg of RNA per ml. Prior measurement, calibration of the instrument with suitable buffer used for RNA elution was applied. The ratio of absorbance at 260 nm to 280 nm (A_{260}/A_{280} ratio) gives an estimate of RNA purity with respect to protein contaminations. Pure RNA has an A_{260}/A_{280} ratio of 1.9-2.1.

RNA integrity

Prior performance of the *in vitro* reverse transcription steps, the integrity of the isolated total RNA was tested by agarose-formaldehyde denaturing gel electrophoresis or for more rapid and precise characterization of the isolated total RNA, Agilent 2100 bioanalyzer (Agilent Technologies, Palo Alto, CA, USA) was used. According to the bioanalyzer, the eukaryotic total RNA is classified based on a numbering system from 1 to 10 as 1 corresponds to the most degraded RNA and 10 to the most intact RNA. The RNA integrity number (RIN) of the analyzed samples, determined by this technique was between 6.4 and 9.6.

Complementary DNA (cDNA) synthesis

First strand cDNA was synthesized from total RNA using oligo (dT) 12-18 primers and SuperScript II Reverse Transcriptase according to the provider. cDNA was diluted with appropriate volume of RNase-free water and stored at -20°C till further processing.

Genomic DNA (gDNA) extraction

The cells or tissues were resuspended in 500 µl of lysis buffer. After addition of 25 µl, Proteinase K (10 mg/ml), the mixture was incubated overnight at 55°C. The proteins were precipitated by adding 300 µl, 5M NaCl followed by vigorously mixing of the suspension. After centrifugation at 15 800 g, 400 µl isopropanol was added to the supernatant and DNA was precipitated. The DNA pellet was washed with 70% ethanol, dried at room temperature and then eluted in appropriate volume of TE buffer for 30 min at 65°C.

The gDNA concentration was determined by measuring the absorbance at 260 nm ($A_{260} = 50$ ng/µl), while the DNA purity was verified by calculating the ratio A_{260}/A_{280} . Pure DNA has an A_{260}/A_{280} ratio of 1.7-1.9.

Buffers:

Lysis buffer – 50 mM tris-HCl, pH 8, 100 mM EDTA, 100 mM NaCl, 1% SDS

TE buffer – 10 mM Tris-HCl, pH 8, 1 mM EDTA, pH 8

Design of primers

Altogether, 12 gene of interest (GOI) and 2 reference genes were analyzed. However, it should be considered the fact that besides the lack of DEFA 2 gene, the sequence of the genes encoding the neutrophilic defensins, i.e. DEFA 1 and DEFA 3, differ by only 2 nucleotide substitutions (1 leading to a coding difference and 1 in the 3'UTR) (108). Therefore only one primer pair was used for detection of DEFA 1 and DEFA 3 genes.

All primer pairs were designed, using freely available Primer3 software, version 0.4.0, (<http://frodo.wi.mit.edu/>). To assess the primer specificity, basic local alignment search tools (BLAST, <http://blast.ncbi.nlm.nih.gov/Blast.cgi>) were applied. Notably, the primers were designed to amplify across an intron/exon boundary or in two different exons, separated by a very big intron, thereby preventing amplification of residual genomic DNA. Information about the mRNA reference sequences of the protein-coding genes, their exon boundaries and the size of the introns in between was accessible at: <http://genome.ucsc.edu/cgi-bin/hgGateway>. The amplicon size varied within the range 94 bp to 202 bp. The sequences of the used primer pairs are shown in Table 3-5, page 22. For further verification of the primer specificity, the products amplified by defensin's primer pairs were sequenced.

Verification of the primer specificity

To verify the primer specificity and to ensure that the designed primer pairs were not amplifying additional products in the presence of genomic DNA (gDNA), *in-silico* PCR was implemented (<http://genome.ucsc.edu/cgi-bin/hgPcr?command=start> or <http://www.ncbi.nlm.nih.gov/sutils/e-pcr/>). In addition to this computational procedure, an experimental approach, such as standard polymerase chain reaction (PCR), was carried out. For generation of the amplicon, reflecting the relative abundance of certain target gene, the gene specific primers were tested in a presence of various templates, i.e gDNA and complementary DNA (cDNA). PCR reactions were carried out using Go Tag DNA Polymerase. In addition, to the reaction mixture were added: 5x Go Tag Reaction buffer, 200 μ M dNTRs mix (equimolar, dGTP, dCTP, dATP and dTTP), gene specific primers (10 μ M) and DNA template (either gDNA or cDNA).

Control reaction (NTC) that contains all essential components of the amplification reaction except the template enables detection of contaminations. Reactions were run on Mastercycler gradient (Eppendorf) with the following thermal conditions: 95°C for 2 min followed by 40 cycles of 95°C for 30 sec, 60°C for 30 sec and 72°C for 30 sec and as a last step, final extension was run at 72°C for 10 min. On Figure 3-1, representative examples are shown. If the primers are specific, amplicon with correct size should be observed only in case where cDNA was used as a DNA template.

Buffers:

5x Go Tag Reaction buffer – 50 mM Tris-HCl, pH 9; 50 mM NaCl and 5 mM MgCl₂

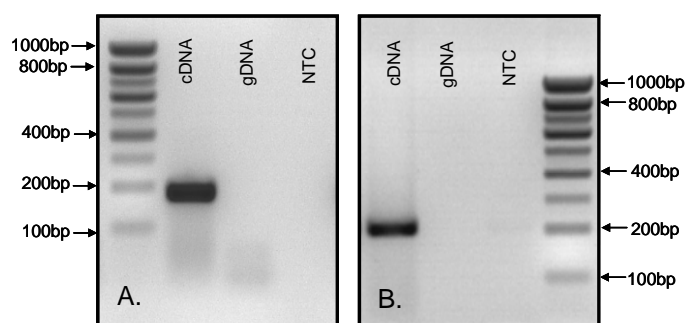


Figure 3-1 Verification of primer specificity for (A.) Defensin, alpha 6 gene (*DEFA 6*) and (B.) Defensin alpha 1-3 genes (*DEFA 1-3*). The expected amplicon size is 198 bp and 222 bp, respectively. In order to verify the lack of contaminations, non-template control reaction (NTC) was run in parallel.

Table 3-5 Sequence of oligonucleotide primers used for PCR amplification and product size predicted for sample cDNA. All primers are listed as 5' to 3'.

gene name	gene annotation	Ref Seq ID	sequence	amplicon size
<i>HSP90β</i>	Heat shock protein 90 kDa alpha (cytosolic), class B member 1	NM_007355.2	CGTTGCTCACTATTACGTATAATCCT CGAATCTTGTCCAAGGCATC	202 bp
<i>PKM2</i>	Pyruvate kinase, muscle	NM_002654.3	TCCGGATCTCTTCGTCCTTG TGGGTCTGAATGAAAGGCAGT	110 bp
<i>PLK1</i>	Polo-like kinase 1	NM_005030.3	ACCAAGGTTTTCGATTGCTC CTCAGGCAGGGGGTCTC	94 bp
<i>COX-2</i>	Cyclooxygenase- 2	NM_000963	CGCTCAGCCATACAGCAA GAATCCTGTCCGGGTACAATC	110 bp
<i>eIF4E</i>	Eukaryotic translation initiation factor 4E	NM_001968	ATCCCCCGACTACAGAAAGAG AGATCAGCCGCGAGGTTTG	153 bp
<i>OPN</i>	Secreted phosphoprotein 1 (osteopontin, bone sialoprotein 1 early T-lymphocyte activation 1)	NM_001040058	TGGAAGTTCTGAGGAAAAGCAG GGCTTTCGTTGGACTTACTTG	173 bp
<i>USP28</i>	Ubiquitin specific protease 28	NM_020886	GAGAAAATTACAAATAAGCTGGAA CCTGAGCCATATTCACATACCT	167 bp
<i>HAT1</i>	Histone acetyltransferase 1	NM_003642	CCTACAGTTCTTGATATTACAGCGG TGTGCCTCTATCGCCATATCTT	152 bp
<i>DDX6</i>	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 6, (RNA helicase, 54kD)	NM_004397	GGAAAAGCCATCTCCTATTTCAG TCAGGTCTAGCCGTTCAAGTAAG	134 bp
<i>DEFA 1-3</i>	Defensin, alpha 1 to 3	NM_005218.3 NM_005217.2	CCTGCCTAGCTAGAGGATCTGTG TGTTTTTCCTTGAGCCTGGA	222 bp
<i>DEFA 6</i>	Defensin, alpha 6, Paneth cell-specific	NM_001926.2	CTCAAGTCTTAGAGCTTTGGGCT GGACACACGACAGTTTCCTTC	198 bp
<i>β-actin</i>	Beta actin	NM_001101	AGAGCCTCGCCTTTGCCGAT CCCACGATGGAGGGGAAGAC	160 bp
<i>GUS</i>	Glucuronidase, beta	NM_000181	TGCAGGTGATGGAAGAAAGTG TTGCTCACAAGGTCACAGG	172 bp

Quantitative Real - Time PCR (qRT-PCR) conditions

Aliquots of 2 μ L cDNA was mixed with PCR master mix iQTM SYBR[®] Green Supermix (2 x SYBR Green I, iTaq DNA polymerase, reaction buffer, deoxynucleotide triphosphate mix, 10 mM MgCl₂, 20 nM fluorescein, and stabilizers), and the gene-specific primers in a final volume of 25 μ L. For one specific sample the assay was performed in duplicate or triplicate. The enzyme was activated after an initial two minutes denaturation step at 95°C and 40 cycles of 30 sec at 94°C, 30 sec at 60°C and 30 sec at 72°C. Amplification was performed in an iCycler thermal cycling instrument (iCycler iQ[®] Real-Time PCR Detection System; BioRad) in 96-well microtiter plates. For each experimental run, a control reaction that contains all essential components of the amplification reaction except template was included. This enables detection of contaminations and presence of primer-dimer formation.

PCR reaction efficiency

The slope of a standard curve provides an indication of the efficiency of the qRT- PCR. Upon analysis, its slope should be -3.322 for 100% PCR efficiency. The standard curve was generated using a dilution series of five different concentrations of cDNA, measured in triplicate. The standards were giving a slope between - 3.1 and - 3.4 which was acceptable for accurate quantification. The fluorescent signal from each PCR reaction was collected as the peak-normalized values plotted versus the cycle numbers.

Reactions were characterized by comparing Ct values. The Ct value is a unitless value defined as the fractional cycle number at which the target fluorescent signal passes a fixed threshold above the baseline, when it is always located within the linear phase of amplification. The higher the initial copy number of cDNA for a certain gene target is the lower Ct value this gene has.

3.2.4. Protein expression analysis.**3.2.4.1. Protein extraction****Preparation of lysates from tissue samples**

The frozen colon stripes, stored at -80°C and ready to be used were thawed in prechilled homogenization RIPA buffer. In addition to 1 mM Na₃VO₄ and 1 mM PMSF, protease inhibitor cocktail (Roche) was added to the buffer in order to protect the lysate from the action of proteases. Cellular debris were removed by centrifugation at 15 800 g for 7 min, at 4°C. The supernatant was carefully removed to a fresh, ice-cold tube.

The protein concentration was determined by applying the standard Bradford protein assay (Roti[®]- Quant), where calf serum albumin was used for calibration protein. The sample was ready to be applied for further analysis.

Buffers:

RIPA buffer – 150 mM NaCl, 1.0 % NP-40, 0.5 % Na deoxycholate, 0.1 % SDS, 50 mM Tris, pH 8

3.2.4.2. SDS PAGE and Western blot, quantification by ECL

In order to analyze the protein content, the tissue and/or cell extracts were subjected to SDS PAGE (sodium dodecyl sulfate polyacrylamide gel electrophoresis) and thus separated. The percentage of the gel casted was dependent on the size of the analyzed protein. The smaller the size of the protein of interest, the higher the percentage of the gel is and vice versa. The protein extracts, corresponding to 15 to 80 µg of the total protein amount were boiled in 1 x Laemmli buffer for 5 min at 95°C. Briefly, by 2 min centrifugation step at 15 800 g, the denatured protein extracts were clarified. Once prepared, the samples were ready to be loaded into the wells of the gel (ranged from 10% to 12%). Each gel was run in running buffer with constant voltage of 40 V until the protein went out from the stacking gel, then the voltage was increased to 60 or 80 V for the separating gel. Electrophoresis time depended on the protein size and it was carried on for at least 90 min. Pre-stained, high-range (10-250 kDa) molecular weight standard (PageRuler Prestained Protein Ladder) was used to determine the molecular weight of the proteins. Semi-dry transfer (transblot, Bio Rad) was applied to transfer the separated by SDS PAGE proteins onto polyvinylidene difluoride (PVDF) or nitrocellulose membrane using transfer buffer. Constant voltage of 14V was applied as the transfer time depended on the size of the protein under investigation, which range from 60 to 90 min. The bound proteins were visualized by Ponceau S solution. To block the unspecific binding sites, the membrane was soaked in 5% non fat dried milk, dissolved in PBS-Tween 0.05% for 1 hour at RT and then was probed with primary antibody (determined dilution, refer to Table 3-2, page 17). Incubation, preferably overnight at 4°C was applied. Next, followed by 3x washing with PBS-Tween 0.05% or 0.1% solution, the membrane was incubated with the respective horseradish peroxidase-labeled secondary antibody for 1 hour at RT. Finally, after the last triple rinsing of the membrane, the proteins were visualized by using chemiluminiscent ECL or ECL Plus™ reagents and exposed to Amersham ECL-Hyperfilm.

If necessary to be re-probed with another antibody, the membranes were first washed twice with PBS-Tween 0.1% and then submerged in stripping buffer, thereafter incubated at 50°C for 30 min with occasional agitation. Followed by extensive washing with PBS-Tween 0.1%, and blocking (blocking buffer), the membrane was ready to be used for immunodetection with different antibodies as described above.

Buffers:

1x Laemmli buffer – 10% Glycerol, 62.5 mM Tris HCl, pH 7.5, 2% SDS, 50 mM DTT and Bromphenol blue

Separating gel – 1.5 M Tris, pH 8.8, 0.4% SDS

Stacking gel – 0.5 M Tris, pH 6.8, 0.4% SDS

10x Running buffer – 0.25 M Tris; 14.4% Glycine (w/v)

1x Running buffer – 100 ml 10 X Running buffer, 5 ml, 20% SDS. Fill up to 1000 ml with Millipore-Q Water

10x Transfer buffer – 250 mM Tris (Base); 1.92 M Glycine

1x Transfer buffer – 100 ml, 10x Transfer buffer and 200 ml , 100% Methanol. Fill up to 1000ml with Millipore-Q Water

10x Phosphate-buffered Saline (PBS) – 137 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, 2 mM KH₂PO₄, pH 7.4.

Stripping buffer – 100mM β-mercaptoethanol, 2% SDS, 62.5 mM Tris-HCl pH 6.7

Blocking buffer – 5% non-fat dried milk in PBS – Tween 0.05%

3.2.4.3. Quantification of the chemiluminescent signal by densitometry and evaluation of the Western blot data

Densitometric estimation of the blots was performed using Quantity One quantification software, version 4.5 (Bio Rad). It measures the signal intensity derived from the digital data object. To be visible and quantifiable, the intensity of the clustered pixels from the data object should be higher than the intensity of the pixels that make up the background of the image. The densitometric value used for evaluation is called “Volume” and it is equal to the sum of the intensities of the pixels within the determined volume boundary multiply by the pixel area. Further correction of the value was performed by the followed background subtraction.

The intensity of the target protein was normalized to the signal intensity of the reference protein, i.e β-actin. The last functioned as an internal, loading control.

3.2.5. Data mining, evaluation techniques and statistical methods

3.2.5.1. Data mining

Expression data for the genes of interest as well as the classification of tissues have been obtained via the "Virtual Northern" function of NIH's database dbEST (109-111). Recent short descriptions are available in references (112-114). This database searches in literature for data from DNA or oligonucleotide chip experiments, normalizes them and lists the results gene by gene for a set of tissues. The data can be obtained for each gene via <http://cgap.nci.nih.gov/Genes/GeneFinder>. Presently, the database contains information on some 4 million genes or EST expression data.

3.2.5.2. Principal Component Analysis (PCA)-the idea behind it

In the present work a combination of 12 genes of interest (GOI) as well as 2 reference genes was analyzed in normal and tumor colon tissues. In order to find out which of these selected genes have a high diagnostic power and which ones are of less importance for distinguishing normal from tumor tissue, the statistical technique “Principal Component Analysis” (PCA) was used. PCA is a mathematical procedure that transforms a number of (possibly) correlated variables into a smaller number of essential variables called Principal Components (PC). The description of the algorithm and the applied terminology can be seen below.

To demonstrate how PCA works, an artificial example will be described. Assume that a set of four genes (variables) was used to distinguish tumor from normal tissue obtained from three patients. The putative expression measured by qRT-PCR for each gene (variable) across the two types of tissues was presented as threshold cycle (Ct) values. These expression values were not further normalized and raw data was used for evaluation (Table 3-6). The resulting data showed that Gene C and Gene D had very stable expression values among the analyzed tissue samples, which lead to the assumption, that these two genes would not have any impact in the discrimination between the analyzed normal and tumor tissues.

Table 3-6 Expression of Gene A, B, C and D (columns) measured by qRT-PCR in normal (N) and tumor (T) tissue for three patients (rows). The numbers indicate the different patients. Raw expression data are presented as a number of threshold cycles (Ct). No additional normalization of the expression was performed.

Type of tissue	Gene A	Gene B	Gene C	Gene D
N1	26	25	26	21
N2	27	25	25	23
N3	25	27	27	22
T1	29	28	26	21
T2	30	28	25	23
T3	28	30	27	22

Next, the obtained qRT-PCR data were applied to the PCA software and the outcome of the analysis is shown in Figure 3-2, page 27. It should be noted that due to the PCA performance the scale of the scatter presentation was changed.

The assay disclosed that discrimination between “normal” and “tumor” tissue population can be achieved and for this positive outcome, the first principal component (PC1) had the most prominent impact since it showed the largest variance in the analyzed data. PC1 accounted for more than 63% of the variance in the data and this was proved by the eigenvalue (detailed description, below), which had the highest value (5.5), compared to the eigenvalues of the rest of the principal components (Figure 3-2, B). Next, in order to verify the most important variables (genes), responsible for the prominent role of PC1, its “component loading” was evaluated (Figure 3-2, C). The observation revealed that Gene A and Gene B were the genes with the heaviest impact that enable discrimination between the two types of tissues. They both had almost equal loadings in this component. In contrast, Gene C and Gene D did not show any discrimination impact.

The finding that the discrimination power of the analyzed genes was not equal could also be clearly observed and verified when the expressions of two different genes were plotted in combination. Figure 3-3, page 28 summarized the assembled 4 different combinations of genes. While the combinations of genes, where the most prominent targets, such as Gene A and Gene B were implicated, can distinguish normal from tumor tissue (Figure 3-3, A, B and C), the combination of genes, which did not show any alterations in the expression

between the different types of tissues, as Gene C and Gene D, was unable to discriminate between normal and tumor tissue (Figure 3-3, D, page 28).

Figure 3-2 (A.) Scatter of the first 2 Principal Components (PC) computed after performance of PCA with four variables, e.g. Gene A, Gene B, Gene C and Gene D. The raw data were transformed to four Principal Components (PCs) where the first one has the biggest impact for the separation between the normal (green) and tumor (red) tissue populations. In addition; (B.) presents a table where the percentage variance for each of the 4 computed PCs is listed together with the corresponding eigenvalue; (C.) A schematic diagram showing the “component loading” of PC1, which in fact facilitates the user to determine to what degree the different original variables (genes) enter into the most prominent component.

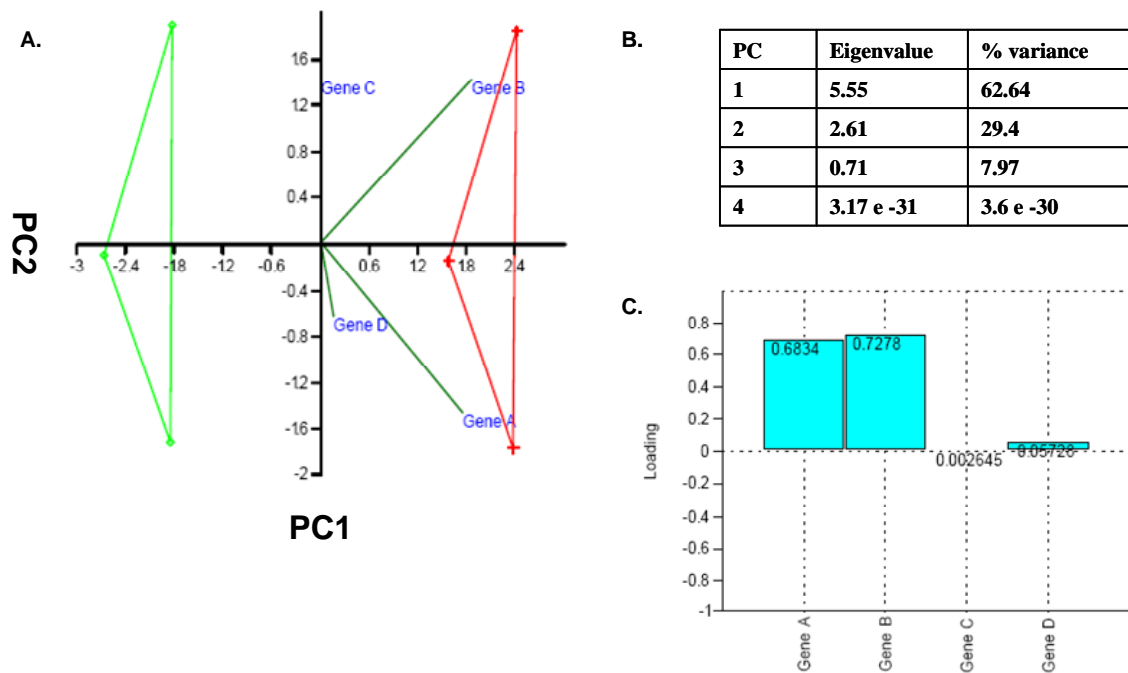
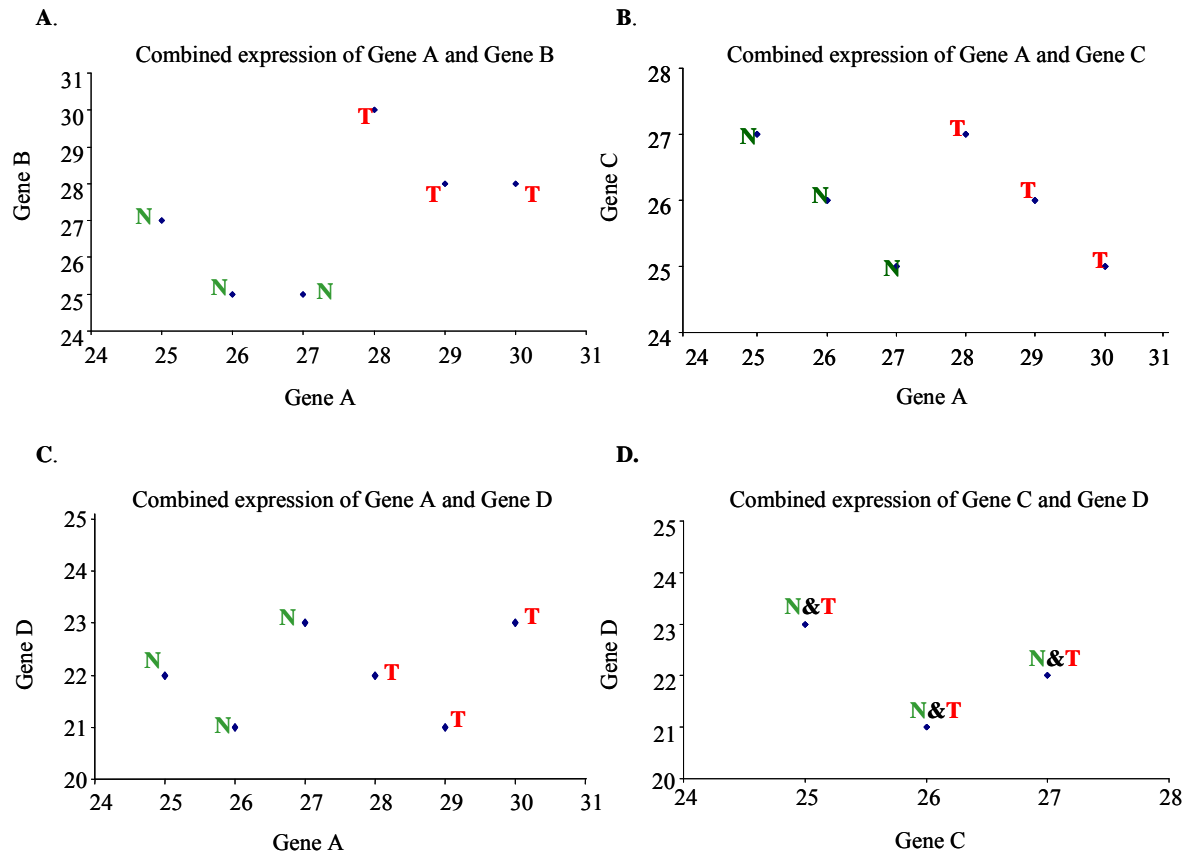


Figure 3-3 Scatter of the pairwise expression of 4 genes examined. (A), (B) and (C) show cases where distinction between normal (N) and tumor (T) tissue is possible; (D) presents the combination of Gene C and Gene D. Their expressions did not differ when N and T tissues were compared (Table 3-6, page 26). Therefore, the combination of these 2 genes is not able to discriminate between the two different tissue types analyzed.



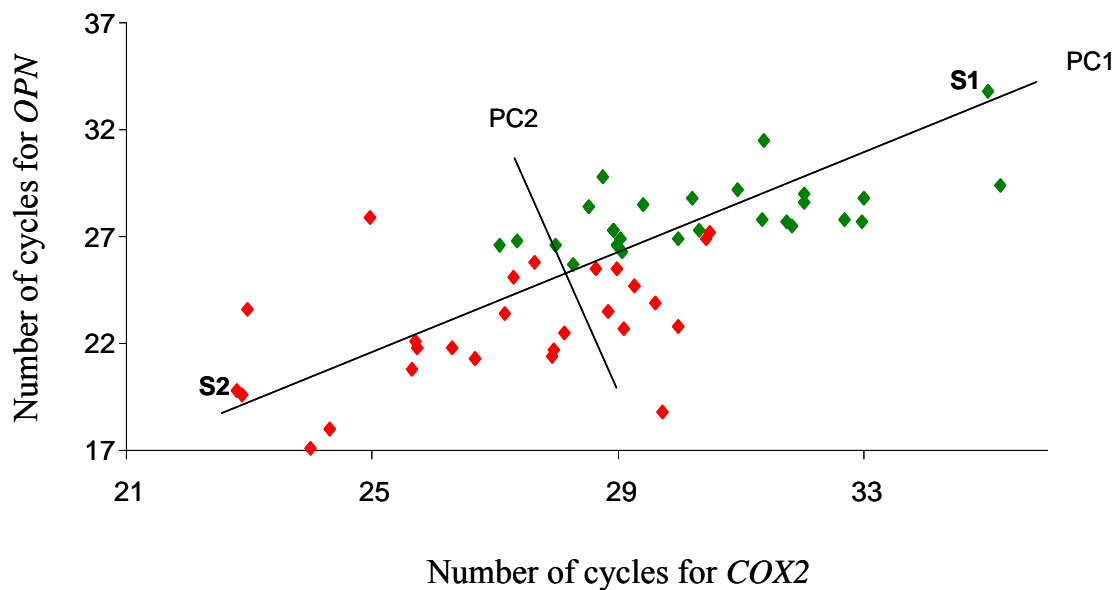
Background of PCA. Description and terminology.

As described above, PCA is a statistical technique that transforms a number of (possibly) correlated variables into a smaller number of essential variables called PCs. For further understanding of the way how this analysis works, here is given an example with actual experimental data, namely gene expression data obtained from paired normal and tumor colon tissue. In this case a *single variable* is the measured expression of a defined gene across different colon tissue types. In a qRT-PCR experiment, the gene expression is represented by Ct values. As it was mentioned above, these measures were not normalized and raw data were applied to the software. It should be taken into account that when multivariate data are evaluated (parallel evaluation of many genes (variables)) not all, but at least some of the variables can often be correlated with each other. Such an example is shown in Figure 3-4, page 29, where two highly correlated variables were plotted, i.e. *COX-2* and *OPN* gene expressions across normal and tumor colon tissue. These two variables

were correlated with each other, since if in one sample the gene expression value of *COX-2* was high, the value of *OPN* variable was high as well and/or vice versa. This presentation closely resembles Figure 3-3 A, page 28. Thus, after performance of PCA, the data of these two variables were transformed into PCs. These PCs are a linear combination of the original variables and they are equal to the number of the original variables.

The PCs are always orthogonal/perpendicular to each other and therefore they are uncorrelated. By definition, the first principal component (PC1) accounts for as much of the variability in the data as possible and every succeeding component accounts for as much of the remaining variability as possible. In the present example, two PCs were calculated, termed PC1 and PC2 where the first one (PC1) possessed the main variance in the analyzed data. Principal component 2 (PC2) was the second component accounted for the remaining variability (115; 116).

Figure 3-4 Scatter plot of two variables. The romb symbol “♦” represents the measured, non-normalized expression of two genes, e.g. *OPN* and *COX-2* in one sample, presented as number of threshold cycles (Ct). The PCA transformed the analyzed variables into two PCs where PC1 gives the major variance in the data. S1 and S2 illustrate respectively the samples with the highest and the lowest relative expression among the different samples and hence, represent the major variance in this data set.



In addition, Figure 3-4 highlights samples S1 and S2, which represent the largest variance in this data set. Every sample point (♦) indicates the non-normalized, raw expression of *OPN* and *COX-2* in one tissue donor as the colored in red points represent the tumor samples and the green one, the normal samples. Thus, S1 and S2 symbolize the donors with the highest and the lowest *OPN* and *COX-2* expression levels among the entire set of investigated variables. Obviously, PC1 was closely passing by S1 and S2, symbolizing the highest

variance in the data, while PC2 was covering the second main variance presented in that data set. When 3 additional variables, for example the expression of *PLK1*, *PKM2* and *USP28* genes were added to the already assessed *OPN* and *COX-2* data, five variables (expression patterns) in total were analyzed. Interestingly, the 3 new variables were highly correlated within each other, but they were not correlated with the first 2 variables (*OPN* and *COX-2*). Therefore, among the data, one can suggest that 2 different groups were differentiated, each consisting of highly correlated variables. After evaluation of an entire set of 5 variables, PCA projected the data onto new variables (PCs) as the first two PCs were those with the most prominent function. Probably these two PCs demonstrate the relationship between the 2 newly determined groups, *OPN* and *COX-2* genes on one hand and *PLK1*, *PKM2* and *USP28* genes on the other. Indeed, via applying PCA some of the components with lesser significance among the investigated data set can be ignored, while the most informative will remain. As a result, in the present example the number of variables can be decreased from 5 to 2, as the latter were those with the most prominent role.

The measure of variance in the data associated with a PC is called eigenvalue. This value can be used to assess the relative importance of the PC. If the PC has a low eigenvalue, then it is contributing little to the explanation of variances in the variables and may be ignored as redundant with more important PCs. As it was mentioned above, the first PC possesses the highest eigenvalue and gives the direction of maximal variance. The second PC gives the direction of the second highest variance orthogonal to the first PC and etc. Thus, the final result of applying PCA is a data compression, which reduces the high dimensionality of the experimentally obtained data to fewer dimensions, with a minimum loss of information. In this way, the most relevant components, or patterns, across the gene expression data which are responsible for the group's similarity or differences will be identified.

For more detailed description, the reader can refer to:

<http://www.chem.agilent.com/cag/bsp/products/gsgx/Downloads/pdf/pca.pdf>

Description of the software used for PCA analysis

The software used for the PCA of the gene expression data is available at: <http://folk.uio.no/ohammer/past/>. The outcome from the PCA can be displayed as a 2D picture, viewed from its most informative viewpoint when representing the first two most prominent PCs. In addition to that, the data analysis package reports the list of all computed PCs as well as the corresponding eigenvalues, which indicate the most significant PCs. The software enables the researchers to verify the “component loadings”, which in fact means to test to what degree the different original variables (given in the original order along the x axis) enter into different components. These “component loadings” are important when the user tries to interpret the “meaning” of the components.

To implement the PCA, raw Ct data obtained by qRT-PCR was utilized. For a detailed description of the Ct value refer to page 23, “PCR Reaction Efficiency”. In order to perform the PCA approach, it should also be considered that the applied raw Ct data must be

acquired using the same starting cDNA concentration. However, this requirement was not fulfilled for each of the analyzed patient's materials. Therefore, prior to application of the software, the raw Ct data from those exceptional cases was undergoing additional adjustment as described below:

Ct is the number of cycles when the fluorescence intensity of the target gene in qRT-PCR reaction passes a fixed threshold value V_t (equation 1)

$$V_t = C_0 * 2^{Ct} \quad (1)$$

where C_0 is an initial cDNA concentration.

Therefore, according to equation 1, the initial concentration is:

$$C_0 = V_t / 2^{Ct} \quad \text{or} \quad C_0 = V_t * 2^{-Ct} \quad (2)$$

From equation 2 follows that:

$$\log_2 (C_0) = \log_2 (V_t) - Ct \quad \text{or} \quad -Ct = \log_2 (C_0) - \log_2 (V_t) \quad (3)$$

Under identical PCR conditions – Ct represents a measure for the logarithmic initial concentration of the amount of cDNA in the reaction (equation 3).

In order to compare the qRT-PCR reactions with different starting cDNA concentrations, a correction factor λ was used for the adjustment. The corrected values are given as:

$$C_1 = C_0 * \lambda. \quad (4)$$

According to the equation 3, the outcome of the qRT-PCR analysis will be:

$$\log_2 (C_1) = \log_2 (V_t) - Ct_1 \quad \text{or} \quad \log_2 (C_0) + \log_2 (\lambda) = \log_2 (V_t) - Ct_1 \quad (5)$$

or

$$\log_2 (C_0) = \log_2 (V_t) - Ct_1 - \log_2 (\lambda) \quad \text{or} \quad -Ct = -Ct_1 - \log_2 (\lambda) \quad (6)$$

Thus, if a reaction on the basis of $C_1 = C_0 * \lambda$ is used instead of C_0 for the determination of the initial concentration, the measured value of $-Ct_1$ has to be corrected to $-Ct_1 - \log_2 (\lambda)$.

3.2.5.3. Relative expression software tool (REST)

REST is a program, available as a free download at <http://rest.gene-quantification.info/>. The software was used for relative quantification of a gene expression data, generated by qRT-PCR. REST compares two groups, with up to 16 data points (genes) in the sample group versus 16 data points (genes) in the control group, which allows the use of more than just one house keeping gene (HKG) for evaluation of data. In this particular study, the cohort of normal tissue samples was always regarded as “control group”. The rest of the examined tissue-populations, e.g. normal near tumor, adenoma and/or tumor were considered as “sample groups”.

The data presented to the software was the mean Ct value obtained from samples amplified in duplicate or triplicate by the PCR approach. For a detailed description of Ct, refer to section “PCR Reaction Efficiency”, page 23.

Equation 7 shows the mathematical model used for computation of the expression ratio (R) between the relative expression of the gene of interest (GOI) in the control and the sample groups obtained by qRT-PCR.

$$R = \frac{(E_{GOI})^{\Delta Ct_{GOI}(\text{control-sample})}}{(E_{ref})^{\Delta Ct_{ref}(\text{control-sample})}} \quad (7)$$

where E is the PCR efficiency and Ct represents the mean of Ct, provided by the iCycler, for the GOI and for the reference (ref) gene.

In this work, the PCR efficiency of all applied primer pairs was comparatively similar, and also was high enough to be considered as 100% (which means that with each PCR cycle, 2 copies of mRNA are amplified). Therefore, an optimal primer efficiency of 2 (E=2.0) was assumed. By REST, comparison between the relative expressions in the control and sample groups can be achieved (117). In order to test whether the expression alterations between two analyzed groups are significant, a nonparametric test was automatically carried out by the software. For comprehensive description refer to (117).

To compute the relative expression of a single gene of interest, its mRNA expression was normalized to that of the reference mRNA (equation 8):

$$2^{(Ct_{\text{reference}} - Ct_{\text{gene of interest}})} \quad (8)$$

3.2.5.4. Statistical methods

In order to perform the appropriate statistical tests, the data was exported to Prism Software, version 4 (Graph Pad, San Diego, USA).

If the data followed Gaussian bell-shaped distribution, parametric tests were applied. However, often a mix of Gaussian and non-Gaussian distributed data set was compared and therefore, the use of parametric test was not appropriate. Thus, non parametric tests were applied. These tests make no assumption about the data-distribution, but rather rank the values from low to high and therefore assess the distribution of the ranks.

The following tests were applied:

To compare two paired, matched groups, Wilcoxon paired test was performed.

Mann-Whitney test was used in order to evaluate the means of the relative mRNA expression in response to gender of the patients.

To compare the means among three or more unpaired groups, Kruskal-Wallis test followed by Dunn's post test was applied.

Pearson (for Gaussian distributed data) or Spearman (for non-Gaussian distributed data) correlation analyses were used to examine how well two variables vary together.

To explore whether the age of the patients has any effect on the expression level of the analyzed gene, linear regression analysis was applied.

Comparisons among three or more different cell lines were performed using One-way ANOVA (analysis of variances) for normal distributed values. It was not appropriate to apply the non parametric Kruskal-Wallis test due to the small sample size (three values per group).

Non-paired, Mann-Whitney and Kruskal-Wallis test

These are non parametric tests that compare two unpaired groups. The whole set of applied values are ranked from low to high, without paying attention to which of the analyzed group each value belong. The smallest number gets rank of 1, while the largest one gets a rank of N, where N is the total number of values in the examined groups. Next, the ranks from each group are summed up and if the sums are very different, than the p-value will be very small. The latter answer the question: If the populations really have the same median, what is the chance that random sampling would result in a sum of ranks as far apart as observed in this experiment?

If the p-value is small, the idea that the difference is due to random sampling can be rejected, rather can be concluded that the populations have different medians (detailed description “Statistics Guide” provided by Prism SOFTWARE, page: 53-54 and page: 71-73).

Paired, matched Wilcoxon test

It is nonparametric test that compares two paired groups. The test calculates the difference between paired measurements for each analyzed subject. For example, this test was used to compare the differences in the relative gene expression between paired tissue samples, derived from one single patient. Basically, the absolute values corresponding to the computed differences between each set of pairs are ranked from low to high. The ranks from each analyzed group are summed up and the sums are compared. If the sums of the ranks are very different, the p-value will be small. (page: 54-56, Statistical guide, Prism).

One-way ANOVA

In brief, One-way ANOVA works by comparing the differences in the means among three or more unmatched groups (variables) when the data is categorized in a one-way. For more information about how this is accomplished, refer to:

http://www.analystsoft.com/en/products/statplus/content/help/src/analysis_analysis_of_variance_one_way_anova.html

When the p-value is small, than the null hypothesis of no differences between the means is rejected and the alternative hypothesis that the means are different from each other is accepted.

Furthermore, in order to compare all pairs of group means, Tukey's post test was selected. Thus, one can evaluate whether the differences between two paired groups are significant.

Linear regression and Pearson/Spearman correlation tests

While the linear regression finds the line that best predicts Y from X, the correlation quantifies how well X and Y vary together. Moreover, the correlation analysis only makes sense when both X and Y variables are measurable outcomes. If one of the variables is controlled, than linear regression should be applied.

For extensive description one can refer to "Statistics Guide", page 92 and "Regression book", provided by Prism SOFTWARE.

4. Results

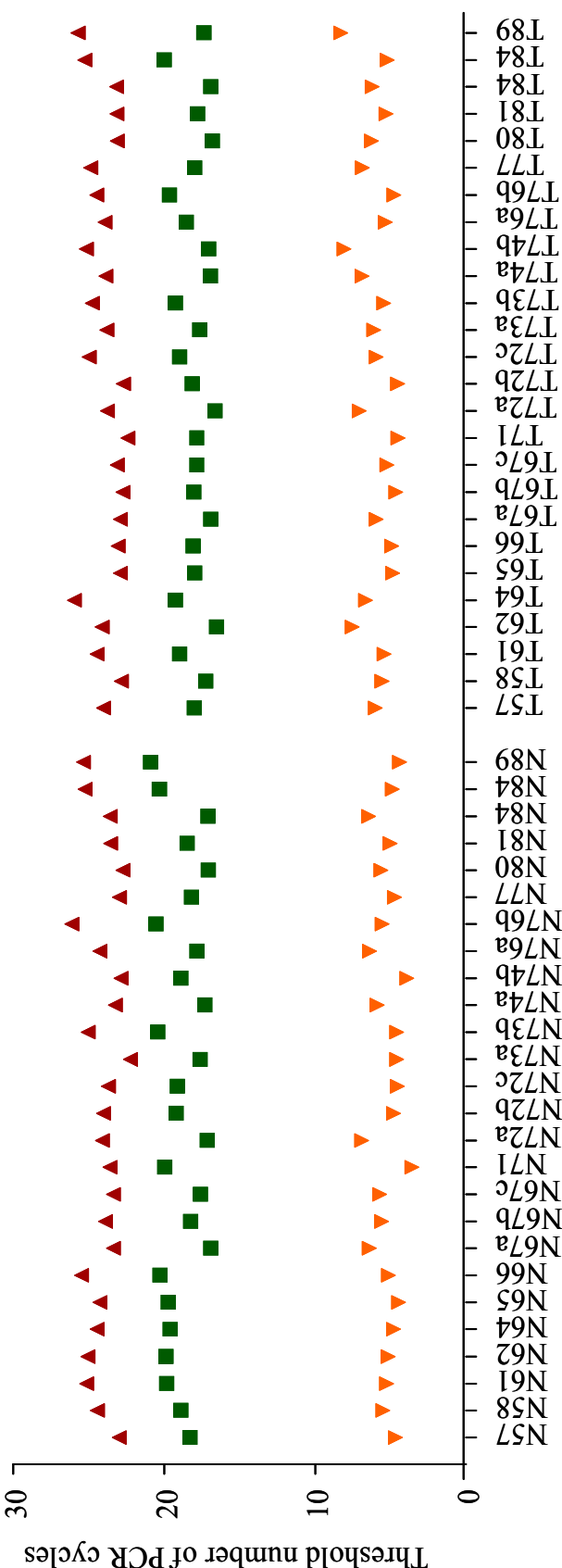
4.1. Quantitation of the housekeeping gene (HKG) expression in normal, normal near tumor, adenoma and tumor colon tissue across patients with colorectal cancer (CRC)

The quantitative Real-Time PCR (qRT-PCR) analysis is a commonly used method to measure and compare the messenger RNA (mRNA) expression levels (transcript abundance) among different tissues and/or disease states. For accurate performance of the assay, several parameters such as RNA integrity, efficiency of cDNA synthesis, initial sample amount need to be considered. It is essentially impossible to obtain the same starting amount of the material under investigation. Therefore, for accurate quantitation of mRNA levels in any tissue samples, it is necessary to apply normalization with endogenous standards, mainly housekeeping genes (HKGs). The proper choice of HKGs highly depends on the tissue and the cell types under investigation. In theory these HKGs, which are responsible for the basic cell survival, should be constitutively expressed in all cell types, without being affected by any human diseases. One of the commonly used endogenous standards, also employed in our investigation, is the β -actin gene. It is ubiquitously expressed in all cell types and additionally it was showing stable expression in colon carcinoma (118). However, many studies have shown differences in the expression profiles of endogenous reference genes between different tissues, pathological states and/or different experimental conditions (118-121). Apart from these observations, previous studies have already reported the relatively stable expression of the gene encoding the lysosomal enzyme β -Glucuronidase (GUS) in normal tissue compared to tumor colorectal tissue (118). Therefore, for more accurate and reliable normalization of our experimental data, the expression of *GUS* was selected as a standard, in addition to the expression of the widely used β -actin gene.

The selected reference genes belong to different functional classes. While the β -actin is a cytoskeletal protein, essential for the structure and kinetics of the cytoskeleton, the GUS is a glycoprotein which functions as an exoglycosidase in the lysosomes. This functional independence is important, because selection of genes that share common biochemical pathways can have influence on the assay, due to the fact that their expression patterns can be co-regulated.

Prior to the performance of the comparative qRT-PCR analysis, the PCR efficiency of each primer pair designed to amplify certain HKG as well as the quality of each RNA sample was determined to ensure the correctness of the assay. Furthermore, in order to validate the expression stability of the HKG across different tissue types, raw (non-normalized) expression data obtained from qRT-PCR assay were plotted in Figure 4-1, page 36. This figure presents the cycle number at which the threshold fluorescence was obtained for the both calibration genes in paired normal and tumor tissues, across 26 donors.

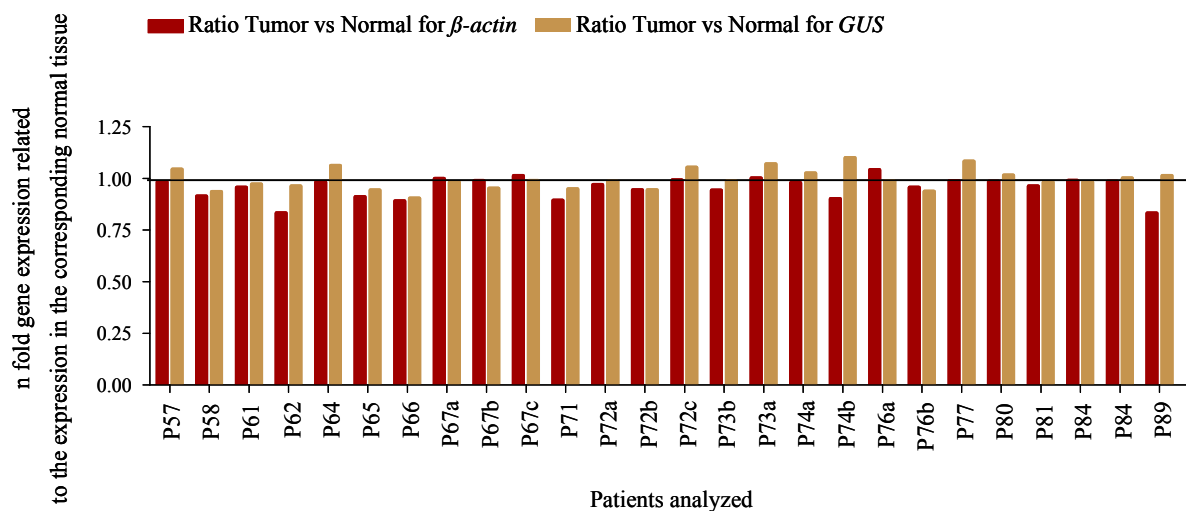
Figure 4-1 Threshold number of PCR cycles required for finding a first florescence signal above background of *GUS* (▲), *β-actin* (■) and difference thereof (▼) for 26 donors analyzed. The left part on the graph represents the expression of *GUS* and *β-actin* in normal tissue (N) and the right part shows the respective expression pattern in tumor tissue (T), derived from patients having colon surgery. Every single patient is denoted with capital letter, corresponding to the tissue type analyzed and number which indicates the age of the patient tested. The small types (letters) describe different patients with the same age.



The left part of the graph presents the threshold cycles in normal (N) tissue samples and the right- in the tumor (T) tissue samples. This way of representation revealed that the both calibration genes need relatively the same number of PCR amplification cycles in order to become detectable, in normal and in tumor colon tissue.

It was also shown distinction between the abundance levels of both HKGs in the cell. The differences ($\Delta(GUS - \beta\text{-actin})$) in the expression between *GUS* and $\beta\text{-actin}$ were approximately 5 - 6 cycles ($5.18 \text{ cycles} \pm 0.79$ for normal tissue and $5.86 \text{ cycles} \pm 1.08$ for cancer tissue), i.e. the absolute expression level of $\beta\text{-actin}$ was 32 to 64 times higher than the expression level of *GUS*. In addition to this verification, another way to test the expression stability of the respective normalization gene was to build-up the ratio of $\beta\text{-actin}$ and/or *GUS* in paired colon tissue samples. Thus, Figure 4-2 exhibits $\beta\text{-actin}$ and *GUS* ratio in matched normal and tumor tissue across 26 individual donors.

Figure 4-2 Comparative HKG expressions as determined by qRT-PCR across 26 donors. Every single bar represents the ratio computed between the expressions of defined HKG in paired non-malignant and malignant colon tissues, derived from single patient, see the legend above the figure. In this ratio, the expression in non- malignant tissue was always dominator. Ratio above or under 1 indicates variation in the HKG expression in colon cancer tissue compared to normal colon tissue. The patients (Ps) are listed according to their age, the small letter discriminate among patients with the same age.



If the computed expression ratio is equal to 1 then the reference expression is quite stable (constant) and not affected by any disease statement. In contrast, any fluctuations of the ratio values (different from 1) will correspond to decrease in the HKG's expression stability. One might expect that, since the cancer is a disease with up-regulated cell proliferation on the one hand and on the other – the cytoskeleton plays a role in the cell migration that $\beta\text{-actin}$ expression would change with increasing disease stages. However, this was not the case, since the obtained result revealed minimal variances in the reference genes expression independent of the type of colon tissue which was investigated.

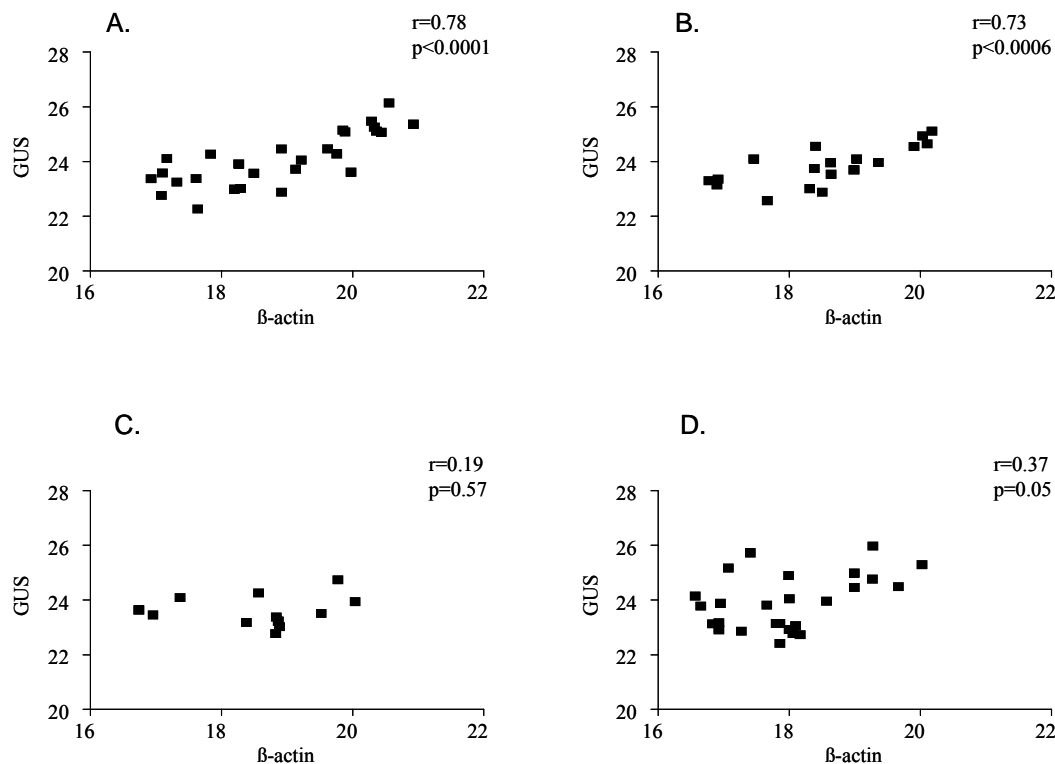
Thus, the analysis of the present study showed that the average fold change in the β -actin expression, in malignant colon tissue, when compared to the corresponding expression in non-malignant (normal) tissue was 0.95 fold \pm 0.05 and 0.99 fold \pm 0.05 for *GUS*, respectively. Furthermore, the HKGs mRNA transcriptional levels were also stable in the rest of the examined tissues, namely normal near tumor (NnrT) and adenoma (A) tissues, when their expression levels were compared to the respective expression in normal tissue. The averaged data is summarized in Table 4-1.

Table 4-1 Averaged fold changes in the HKG expression related to the expression in normal colon. Fold change equal to 1 denotes that there is no change in the HKG expression in certain type of tissue compared to the expression in normal colon tissue.

Type of tissue	Fold changes in β -actin expression	Fold changes in <i>GUS</i> expression
Normal near Tumor vs Normal tissue	0.97 \pm 0.05	0.98 \pm 0.03
Adenoma vs Normal tissue	1 \pm 0.054	0.98 \pm 0.045
Tumor vs Normal tissue	0.95 \pm 0.05	0.99 \pm 0.05

Next, the similarity of the HKG-expression pattern across the entire set of different tissue samples was measured. A metric analysis (Pearson correlation test) was applied. Interestingly, the analysis revealed statistically significant correlation in the expression of β -actin and *GUS* in normal, normal near tumor and tumor colon tissue, while no correlation was found in adenoma colon tissues, see Figure 4-3, page 39. In fact, one reasonable explanation for the lost of correlation between the expression patterns of β -actin and *GUS* in adenoma colon tissue, could be the low number of donors analyzed.

Figure 4-3 Pearson's correlation analysis of the relative HKG expression. The x- and y-axis represent the Ct values of β -actin and *GUS* expression, respectively obtained in Normal (A.), Normal near Tumor (B.), Adenoma (C.) and Tumor (D.) colon tissues. The Pearson correlation (*r*) coefficient represents the similarity in the expression pattern between the two HKGs. *r* equal to 1 reflects perfect correlation between two expression patterns. Statistical significance is denoted by the p-values.

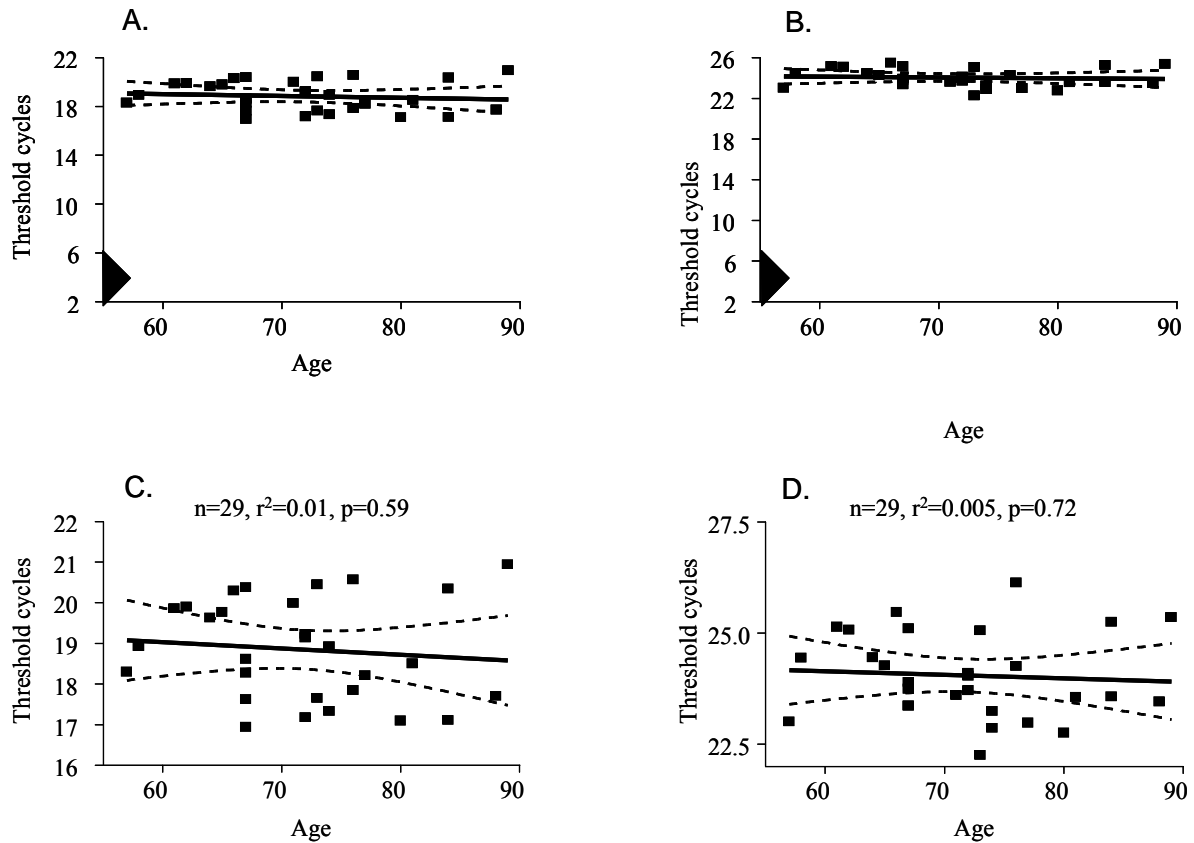


Housekeeping genes are stably expressed at age from 57 to 88 years

The linear regression analysis of the HKGs expressions in normal tissue samples illustrates that there is no dependence on the age (Figure 4-4, page 40). This leads to the conclusion that, at least in the case of β -actin and *GUS* the expression is constant up to high age.

As a further verification of the relatively stable HKG expression, the expression data was categorized according to gender of the patients. No significant differences between the groups were found when Mann-Whitney test was applied (data not shown). The same was found when the HKGs expression in malignant colon tissue was compared with grade and stage of cancer (using Kruskal-Wallis test), data not shown.

Figure 4-4 Linear regression analysis of age of the patients versus the expression of β -actin (A.) and *GUS* (B.); (C.) and (D.) present the same data as in (A.) and (B.), but with highly suppressed zero point on the y-axis. The correlation coefficient (r^2) and the significance level (p) were calculated. The r^2 values close to zero indicate that there is absolutely not age dependence.



Concluding remarks:

Apart from the diverse functional activity, the analyzed HKGs revealed high expression stability. β -actin and *GUS* showed a minimal variance in their expression patterns, independent of the tissue type investigated as well as independent of the difference in their abundance in the cell. Moreover, the Pearson correlation analysis demonstrated that their expression patterns were significantly correlated among the normal, normal near tumor and tumor tissues, derived from patients having colon surgery. In contrast, no correlation was found in adenoma colon tissue. The analysis also showed that the HKG expression is not influenced by age, gender or type of tissue. As a consequence, these observations verify the capability of the selected HKGs to be used for further normalization of the experimental gene expression data.

4.2. No significant differences between single or multi-housekeeping gene normalization, using Relative Expression Software Tool (REST)!

Usually, the outcomes of many expression studies are relying on data, normalized on a single control gene. However, numerous studies reported that HKG expression can vary considerably as a result of different treatments and/or disease stages (121; 122). Due to such expression variability, the normalization to single HKG can lead to relatively large errors in a significant proportion of the tested sample (123). Therefore, multiple HKG normalization can be applied. In the present work, normalization on one (single) and on combination of two HKGs was performed.

In section 4.1, page 35 already showed that *β-actin* and *GUS* were similarly well suited to be used as standard, calibration genes. Moreover, they both are representative of different biochemical pathways, which are not co-regulated and therefore can be speculated that the parallel calibration with both of them can cover better the molecular changes occurring during the colon cancer development.

In order to find out whether the data evaluation will be similarly accurate if the gene expression is calibrated on a single HKG or on a combination of two HKGs, the experimental gene expression data was normalized once to a single HKG and in addition, it was also calibrated to a combination of two HKGs. The test was performed using Relative Expression Software Tool (REST) for group-wise comparison of the relative expression of 12 target genes across paired normal and tumor colon tissue. The data was provided by randomly selected 8 patients from the cohort of 27 analyzed patients in total. For more detailed description of the REST, refer to 3.2.5.3, page 31.

In short, the software performs normalization of the expression for each GOI to single or combination of two HKGs. At the same time the software estimates the alteration of the normalized expression in the “sample group” by comparing it to the respective expressions in the “control group”. Hence, in our experimental data, the expressions of all 12 genes in normal tissue, which represents the “control group” were compared with the corresponding gene expressions in tumor tissue or the so called “sample groups”. The expression ratio, the corresponding fold changes in the expression as well as the outcomes from the automatically performed non-parametric test are listed in Table 4-2, page 42.

The analysis demonstrated that when the same expression data was normalized to a single or to a combination of two reference genes, not considerable differences, in the final outcomes appeared. For instance, in respect to *DDX6*, *HAT1* and *eIF4E* genes, insignificant differences within the range of 1 fold down- or up-regulation in cancer were computed when single and multi-normalization was performed, respectively. In addition, although the same tendency of increased expression in cancer for all analyzed alpha-defensins (*DEFA 1-3* and *DEFA 6*) was detected, only slight deviations in the p-value of the final outcome were observed.

This uncertainty in the results can not be considered as significant error, because any changes in the expression within 1 fold and even more can be related to some deviation in the technical procedure, such as a pipetting variation for example.

Table 4-2 Summary of evaluated gene expression data for 12 ** targets using REST software. The raw data was normalized once on the single reference gene (β -actin) and subsequently on combination of two reference genes (β -actin and *GUS*). Moreover, the relative expressions in the “sample group” were compared with the expressions in the “control group”. Listed is the ratio, the corresponding fold change in the gene expression as well as the p-value derived from the performed statistical analysis. Marked in grey are the significant ones.

Name of target	Normalized on β actin			Normalized on β -actin and GUS		
	Ratio Sample vs Control	Fold changes	p-value	Ratio Sample vs Control	Fold changes	p-value
DDX6	0.86	- 1.16*	0.58	1.09	1.09	0.78
HAT1	0.93	- 1.1*	0.74	1.19	1.19	0.5
HSP90 β	2.6	2.6	0.001	3.33	3.33	0.001
COX-2	15.1	15.1	0.009	19.2	19.2	0.008
OPN	49.32	49.32	0.001	62.62	62.62	0.001
PKM2	1.79	1.79	0.015	2.28	2.28	0.006
PLK1	1.92	1.92	0.18	2.45	2.45	0.07
USP28	1.37	1.37	0.29	1.74	1.74	0.14
eIF4E	0.97	-1.03*	0.89	1.23	1.23	0.2
DEFA 1-3**	7.69	7.69	0.059	9.75	9.75	0.034
DEFA 6	13.93	13.93	0.063	17.69	17.69	0.05

* the value listed as negative corresponds to fold change down-regulation

** *DEFA 1-3* accounts for 2 genes, namely *DEFA 1* and *DEFA 3*. So far no gene has been found for *DEFA 2*.

Concluding remarks:

In addition to the finding that β -actin and GUS were similarly well suited to be used for calibration, the comparative analysis revealed that no considerable changes occurred when the expression data was evaluated to single or to combination of the two reference genes. Based on these observations as well as on the fact that β -actin was also used as a loading (normalization) control for the protein assay, all gene experimental data were solely calibrated to one reference gene, namely β -actin.

4.3. Individuality of the gene expression and changes in cancer

The entire set of targets investigated in this work consists of genes encoding proteins involved in a broad spectrum of cellular processes such as translation initiation, cellular proliferation, mitosis and/or apoptosis (section 1.4, page 7). Moreover, all of them revealed altered expression in a wide variety of cancers, including colon cancer. However, gene expression analysis of fully blown carcinoma does not account for the expression alterations that are critical for the initiation and the development of the cancer. Therefore, qRT-PCR gene expression analysis of all 12 GOI in a patient-resolved manner was performed. Besides the classical combination of paired-normal and tumor tissue, also the matched normal near tumor and adenoma tissue samples were examined.

For most of the investigated genes, the expression data was collected from 29 patients (two of the donors had only non-malignant colon tissues). The exception is the expression data of the alpha defensin molecules (*DEFA 1-3* and *DEFA 6*), which were accumulated from 36 patients in total, as 7 of them have been analyzed in the context of the early PhD work of Franziska Jahns (Department of Nutritional Toxicology, Institute for Nutrition, Friedrich-Schiller-University Jena).

For each of the selected genes, the fold changes in the mRNA expression when two paired groups of tissue were compared, the outcome of the performed Wilcoxon nonparametric test as well as the number of analyzed tissue pairs, were summarized in Table 4-3, page 44. Due to the fact that not all of the investigated objects had the entire set of tissues tested, the number of tissue's pairs differs for any single combination assessed. The proportional changes in the expression for each gene between normal near tumor, adenoma and tumor colon tissue on one hand and normal tissue on another are presented in Figure 4-5, page 46. Each separate graph corresponds to one selected gene, where a single, colored bar displays the fold changes in the mRNA expression in different tissue's types versus the expression in normal tissue.

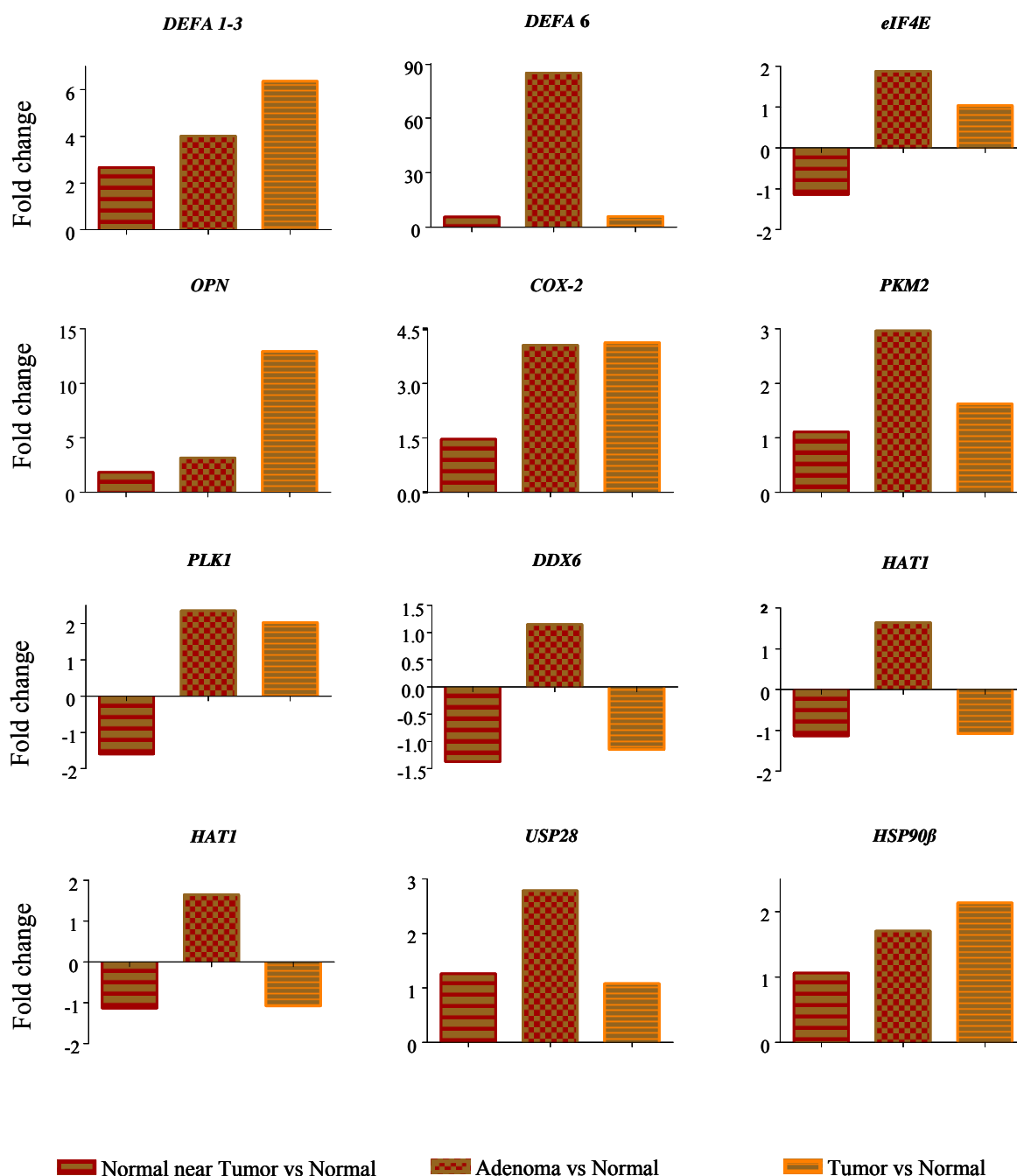
The analysis revealed that in regard to the expression alterations in tumor, *OPN* was the gene showing the highest and significant up-regulation in tumor compared to its expression in normal colon tissue. The *OPN* overexpression in cancer was of around 13 fold. The *OPN* elevation in tumor still remained high when compared with its expression in normal near tumor tissue (8 fold), see Table 4-3, page 44. The second most prominent group of genes in this study was the alpha defensins, namely *DEFA 1-3* and *DEFA 6*, as well as *COX-2*. All of them showed moderate expression increase in fully blown tumors (ranged from 4 to 6 fold). The rest of the analyzed targets revealed relatively low gene expression alterations in malignant colon tissues. Thus, comparatively low (in the range of 2 fold), but significant increase in the expression of the affected, malignant colon tissues was detected for *HSP90 β* , *PKM2* and *PLK1* genes (Table 4-3, page 44). Although not considerable, such constitutively higher expression in tumor cells as compared to their normal counterparts, suggests that they are may be critical for tumor cell growth and/or survival, which has been speculated for the case of HSP90 by Isaacs et al. (124).

Table 4-3 Summary of the mRNA expression alterations when 2 paired groups of tissue are compared. The respective fold changes in the expression are indicated as well as the p-value (Wilcoxon test) which represents the significance in the mRNA expression differences of defined gene when two paired groups are compared. The negative value represents fold down-regulation compared to the group which is denominator in the ratio created; n denotes the number of pairs included in the computation analysis.

	Normal (denominator)	Normal near Tumor (denominator)	Adenoma (denominator)	Tumor (denominator)
PKM2				
Normal (numerator)	X			
Normal near Tumor (numerator)	Fold changes, 1.11 p= 0.22 n=18	X		
Adenoma (numerator)	Fold changes, 2.96 p= 0.0005 n=12	Fold changes, 2.4 p= 0.002 n=10	X	
Tumor (numerator)	Fold changes, 1.62 p<0.0001 n=27	Fold changes, 1.57 p= 0.0003 n=17	Fold changes, -1.54 p= 0.001 n=11	X
DDX6				
Normal (numerator)	X			
Normal near Tumor (numerator)	Fold changes, -1.4 p= 0.035 n=18	X		
Adenoma (numerator)	Fold changes, 1.14 p= 0.8 n=12	Fold changes, 1.7 p= 0.2 n=10	X	
Tumor (numerator)	Fold changes, -1.15 p= 0.2 n=27	Fold changes, 1.1 p= 0.73 n=17	Fold changes, -1.06 p= 0.52 n=11	X
HAT1				
Normal (numerator)	X			
Normal near Tumor (numerator)	Fold changes, -1.13 p= 0.13 n=18	X		
Adenoma (numerator)	Fold changes, 1.64 p= 0.014 n=12	Fold changes, 1.9 p= 0.04 n=10	X	
Tumor (numerator)	Fold changes, -1.08 p= 0.8 n=27	Fold changes, 1.3 p= 0.4 n=17	Fold changes, -1.8 p= 0.08 n=11	X
USP28				
Normal (numerator)	X			
Normal near Tumor (numerator)	Fold changes, 1.3 p= 0.4 n=18	X		
Adenoma (numerator)	Fold changes, 2.8 p= 0.034 n=12	Fold changes, 1.71 p= 0.06 n=10	X	
Tumor (numerator)	Fold changes, 1.08 p= 0.96 n=27	Fold changes, 1.05 p= 0.92 n=17	Fold changes, -1.4 p= 0.12 n=11	X
HSP90B				
Normal (numerator)	X			
Normal near Tumor (numerator)	Fold changes, 1.06 p= 0.69 n=18	X		
Adenoma (numerator)	Fold changes, 1.7 p= 0.002 n=12	Fold changes, 2.7 p= 0.006 n=10	X	
Tumor (numerator)	Fold changes, 2.13 p= 0.0009 n=27	Fold changes, 1.97 p= 0.02 n=17	Fold changes, -1.34 p= 0.64 n=11	X

	Normal (denominator)	Normal near Tumor (denominator)	Adenoma (denominator)	Tumor (denominator)
DEFA 6				
Normal (numerator)	X			
Normal near Tumor (numerator)	Fold changes, 5.4 p= 0.0455 n=19	X		
Adenoma (numerator)	Fold changes, 84.9 p= 0.0003 n=18	Fold changes, 61.3 p= 0.0068 n=11	X	
Tumor (numerator)	Fold changes, 5.8 p= 0.0017 n=33	Fold changes, -2.3 p= 0.45 n=18	Fold changes, -13.82 p= 0.04 n=18	X
DEFA 1-3				
Normal (numerator)	X			
Normal near Tumor (numerator)	Fold changes, 2.67 p= 0.0073 n=19	X		
Adenoma (numerator)	Fold changes, 4.01 p= 0.0007 n=18	Fold changes- 1.33 p= 0.83 n=11	X	
Tumor (numerator)	Fold changes, 6.4 p= 0.021 n=33	Fold changes, 1.15 p= 0.78 n=18	Fold changes, - 1.4 p= 0.71 n=18	X
eIF4E				
Normal (numerator)	X			
Normal near Tumor (numerator)	Fold changes, -1.13 p= 0.62 n=18	X		
Adenoma (numerator)	Fold changes, 1.88 p= 0.03 n= 12	Fold changes, 1.94 p= 0.02 n= 10	X	
Tumor (numerator)	Fold changes, 1.04 p= 0.95 n=27	Fold changes, 1.02 p= 0.56 n=17	Fold changes, -1.65 p= 0.01 n=11	X
OPN				
Normal (numerator)	X			
Normal near Tumor (numerator)	Fold changes, 1.83 P= 0.0023 N=18	X		
Adenoma (numerator)	Fold changes, 3.2 P= 0.04 N=12	Fold changes, 1.34 P= 0.73 N=10	X	
Tumor (numerator)	Fold changes, 12.94 P<0.0001 N=27	Fold changes, 7.9 P= 0.0003 N=17	Fold changes, 11.15 P= 0.001 N=11	X
COX-2				
Normal (numerator)	X			
Normal near Tumor (numerator)	Fold changes, 1.5 p= 0.16 n=18	X		
Adenoma (numerator)	Fold changes, 4.05 p= 0.03 n=12	Fold changes, 3.17 p= 0.08 n=10	X	
Tumor (numerator)	Fold changes, 4.11 p<0.0001 n=27	Fold changes, 3.3 p= 0.0003 n=17	Fold changes, -1.4 p= 0.7 n=11	X
PLK1				
Normal (numerator)	X			
Normal near Tumor (numerator)	Fold changes, -1.6 p= 0.08 n=18	X		
Adenoma (numerator)	Fold changes, 2.4 p= 0.005 n=12	Fold changes, 5.6 p= 0.006 n=10	X	
Tumor (numerator)	Fold changes, 2.01 p= 0.02 n=27	Fold changes, 2.8 p= 0.03 n=17	Fold changes, -1.4 p= 0.06 n=11	X

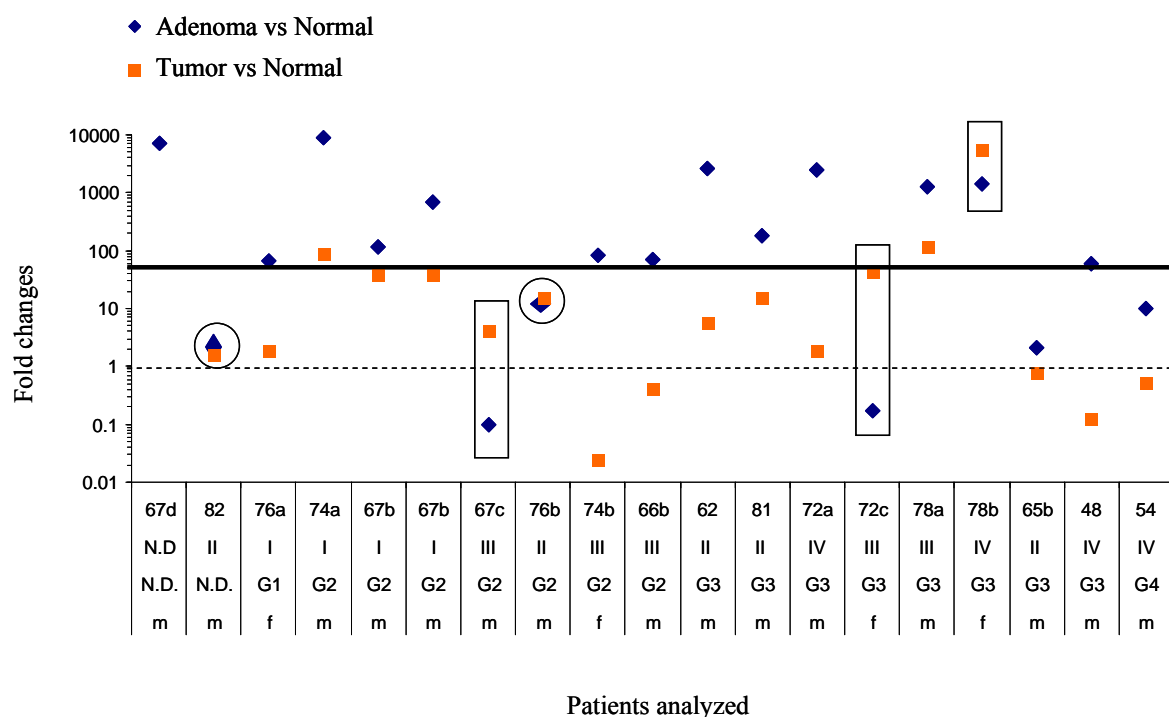
Figure 4-5 mRNA expression alteration in normal near tumor, adenoma and tumor tissues, respectively versus the normal tissue. Each separate graph represents the expression alterations for one single gene and every single bar depicts the fold changes of expression in defined tissue type, in respect to the expression in normal tissue. The fold changes are calculated as medians of the computed ratios between the relative expressions in the both tissue groups which are compared. The expression in normal tissue is always denominator. The outcome of the statistical analysis is presented in Table 4-3, page 44.



Furthermore, more surprising was the dramatic expression burst of *DEFA 6*, with 85 fold in benign adenoma over the expression in normal colon tissue. When compared with normal near tumor tissue, this expression remained very high as well, i.e. 61 fold. Such an increase in one single stage of colon cancer development makes the molecule comparatively strong genetic marker. However, the full potential of these data becomes only evident when it is presented in a patient resolved manner. Thus, Figure 4-6 summarizes the computed *DEFA 6* mRNA expression ratio between adenoma and normal as well as between tumor and normal colon tissue, obtained from each single individual with adenoma tissue. The patient's characteristic such as age, gender, grade and stage of cancer are denoted below the figure. No consistent dependence of *DEFA 6* expression ratios as a function of tumor grade and stage, age or gender has been detected. The ratios in adenoma as well as in carcinoma as compared to normal tissue vary more than 1000 fold indicating the individuality of each patient's cancer.

Figure 4-6 Gene expression regulation of *DEFA 6* mRNA in 18 individuals*. The patients are characterized for their age, gender, grade and stage of cancer. N.D. means that the stage/grade was not determined. The scale of the y axis is logarithmic. None of the patients has simultaneously down-regulated *DEFA 6* in both, adenoma as well as cancer tissue. The dotted horizontal line at 1 indicates that *DEFA 6* would be un-effected. For the first patient on the left, no comparison of adenoma with carcinoma was possible, since the donor had developed only adenoma. Of the 17 other samples, 13 revealed a higher expression in adenoma than in carcinoma, two cases (encircled) showed essentially no change and three objects (in rectangular boxes) had the highest expression in carcinoma.

* One of the donors (67b) had two adenomas, therefore the figure shows 19 tissues.



In addition, in 12/18 cases (67%) *DEFA 6* was overexpressed with more than 60 fold in adenoma (see Figure 4-6, page 47, the black horizontal line is empirical and separates all samples with overexpression of 60 or more fold from the rest of the samples). Most of the tumor tissues showed overexpression with less than 60 fold.

In contrast to *DEFA 6*, rather low were the transcriptional levels of the rest of the analyzed antimicrobial alpha defensins (*DEFA 1-3*) along the normal-andeno-carcinoma axis. These observations were in agreement with the data derived from culture cell lines (section 4.8, page 69). Although low abundant, according to the performed nonparametric test, *DEFA 1-3* mRNA was significantly up-regulated, with 4 and 6.4 fold in adenoma and tumor colon tissue, respectively, when compared to the normal samples. Thus, independent of the common functional activity *DEFA 1-3* appeared to not show expression effects, similar to this which *DEFA 6* disclosed. This data indirectly confirm the fact that the transcription activation of the enteric and neutrophilic defensin-genes is regulated by diverse signaling pathways.

In contrast to the alpha-defensins, the rest of the tested molecules showed gene expression alterations in adenoma tissue, compared to the paired normal colon tissue, in the range from “1.14” fold (for *DDX6*) to “4.1” fold (for *COX-2*), as most of them were significant.

In addition, the analysis revealed that macroscopically normal colon tissue, adjacent to the tumor site has showed though not very substantial, some alterations in the mRNA levels of the genes analyzed, when compared to normal colon tissue. Exception is the significant moderate up-regulation of *DEFA 6* (5.4 fold) in normal near tumor, compared to normal tissues. Those observations confirm that tissues, morphologically similar to normal tissue already have altered gene expression profile (125).

Concluding remarks

Gene expression alterations have been observed in normal near tumor tissues when compared to the respective normal colon tissues. This result indicates that tissue, morphologically appearing as normal, removed not so far away from the tumor site, possesses already modifications on gene expression level. Therefore, their analysis could be of high interest, in order to find molecules critical for the initiation and the development of cancer.

*To our knowledge this is the first study demonstrating the burst of *DEFA 6* gene in human benign adenoma tissue samples, which immediately indicates its potential to be used as a marker for early premalignant stages of colorectal cancer and not solely as a marker for colon cancer detection.*

*The analysis also disclosed that in comparison to the expression of *DEFA 6*, the transcription levels of *DEFA 1-3* mRNA were relatively low across all the investigated tissues. Moreover, *DEFA 1-3* do not show an expression effect, similar to that of *DEFA 6* gene. This data indirectly confirmed the fact that the transcription activation of the both, enteric and neutrophilic defensins is regulated by diverse signaling pathways.*

In addition, the gene with the most prominent elevation in colon cancer is *OPN* with almost 13 fold followed by *DEFA 1-3*, *DEFA 6* and *COX 2*, which exhibited moderate increase in cancer.

4.3.1. Extremely high expression levels of *DEFA 1-3* mRNAs and undetectable *DEFA 6* mRNA transcription in peripheral blood mononuclear cells (PBMCs).

In the present work, several members of the alpha defensin family were analyzed, namely *DEFA 1-3* and *DEFA 6*. Apart from their common functions in innate antimicrobial immunity, they are produced by different cell sources. While *DEFA 1-3* are major component of the azurophilic granules of neutrophils, *DEFA 6* is mainly produced by the Paneth cell of the small intestine. Thus, in order to confirm that the main sources of *DEFA 1-3*- and *DEFA 6*-production are different, besides the analysis of colon tissues with different degree of cancer progression, total RNA isolated from leukocytes (peripheral blood mononuclear cells (PBMCs)) of healthy individuals was tested for transcription of alpha defensins.

The expression levels of *DEFA 1-3* and *DEFA 6* mRNAs were examined in PBMCs derived from eight healthy patients by qRT-PCR. The analysis showed that in contrast to *DEFA 6* mRNA, which was not found to be expressed in the leukocytes, *DEFA 1-3* genes were highly expressed in these samples. Moreover, it was observed that *DEFA 1-3* expression varies drastically among the donors analyzed, (Figure 4-7). From extremely high as it was detected in patients (P) P1, P2, P3 and P5, *DEFA 1-3* expression was converting to very low (P4, P6, P7, P8).

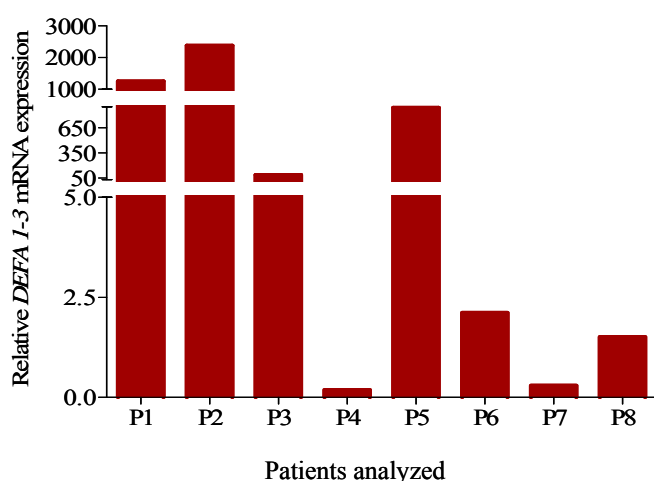


Figure 4-7 Inter-individual variability of *DEFA 1-3* mRNA expression in PMBCs derived from eight healthy patients (P). For better visualization, the scale was segmented.

The observed high inter-individual variation in the *DEFA 1-3* mRNA expression levels among all analyzed PBMC extractions is a consequence of the already reported high copy number variability (CNV) affecting the alpha- and beta-defensins genes clustered on human

chromosome 8p23.1 (96; 126). An additional reason for the disclosed variability could be the demonstrated frequent absence of the *DEFA 3* allele, which has been found missing in significant proportions of the individuals tested (127). However, when compared with the median of the relative expression in colon tissue, the median of *DEFA 1-3* mRNA expressions in PMBCs were tremendously higher (with 7 to 8 orders of magnitude higher), (Figure 4-8). Notice that due to the lack of *DEFA 6* expression in PMBCs, comparative analysis between the expression in leukocytes and colon tissues can not be performed.

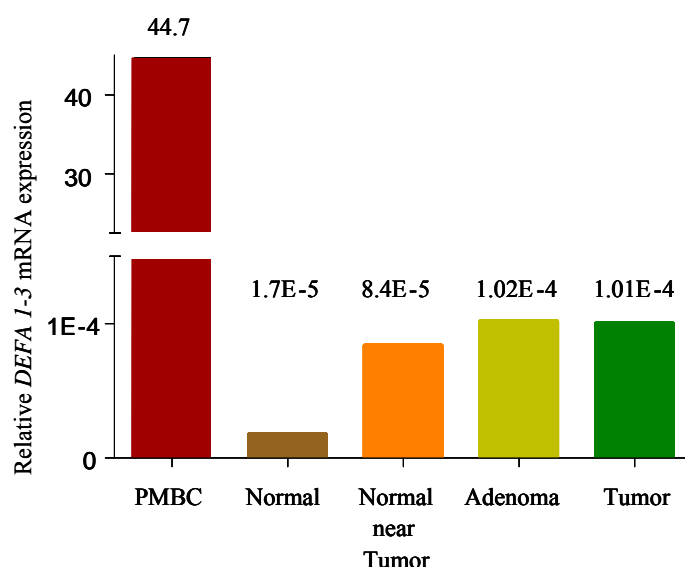


Figure 4-8 Median *DEFA 1-3* mRNA expression in PMBCs from healthy donors as well as in non-, pre-malignant and malignant colon tissues derived from patients who underwent colorectal cancer surgery. On the top of each bar, the medians of the relative expression values are denoted. Due to the high variability in the expression levels, for better visualization the scale was segmented.

4.3.2. Gene expression dependency on the age

In the further investigation, it was analyzed whether and how the mRNA expression levels of the investigated genes correlate with clinicopathological characteristics of the colorectal cancer patients such as age, gender, grade and stage of tumor.

Ageing is the main risk factor associated with cancer development (128). In order to analyze whether the GOI have an expression pattern dependent on the age of the analyzed subjects, Spearman correlation analysis was performed. The expression in normal as well as in cancer colon tissues was evaluated. The analysis showed that none of the investigated genes appeared to be altered in an age-dependent manner. However, slight positive but not significant correlation with the age was found for the expression of *DEFA 6* in normal and tumor colon tissue, see Table 4-4, page 51.

Table 4-4 Spearman correlation analysis exploring the relationship between mRNA expression levels of defined gene of interest and the age of the donors. Spearman correlation coefficients and the respective p-values are listed. None of the investigated genes showed significant correlation in their expression patterns with the age in non-malignant and malignant colon tissue. r range from -1 to +1 as r equal to 1 corresponds to perfect positive correlation.

Gene	Expression in Normal tissue		Expression in Tumor tissue	
	Spearman correlation analysis, p-value	Correlation coefficient, r	Spearman correlation analysis, p-value	Correlation coefficient, r
OPN	0.4	0.16	0.77	0.06
COX-2	0.34	0.19	0.57	-0.12
DDX6	0.84	0.04	0.68	-0.09
DEFA 1-3	0.23	0.21	0.79	0.05
DEFA 6	0.11	0.28	0.15	0.28
eIF4E	0.68	-0.08	0.3	-0.21
HAT1	0.60	-0.10	0.37	-0.18
HSP90B	0.43	0.15	0.33	-0.2
PKM2	0.42	0.15	0.8	-0.05
PLK1	0.57	0.11	0.44	-0.18
USP28	0.32	0.19	0.61	-0.10

4.3.3. Analysis of expression levels in carcinomas according to the tumor grading

The gene expression pattern was examined for dependency on the grade and the stage of the tumor.

Tumor grade and stage are two different tumor classification systems. While the tumor grade categorize the cells in terms of how abnormal they look under the microscope and how quickly the tumor is likely to grow and spread, the determination of staging is based on complex of factors such as location of primary tumor, tumor size, number of tumors and lymph involvement.

In this study, the grade of the entire set of samples derived from colorectal cancer patients was classified according the scale from 1 to 4, namely from Grade 1 (G1) to Grade 4 (G4). The cells of G1 tumors resemble normal cells, which tend to grow and multiply slowly. G1 tumors are generally considered as the least aggressive in behavior, while G3 or G4 tumors do not look like normal cells of the same type and tend to grow rapidly and spread faster than tumors with lower grade.

As mentioned above, the investigated defensin genes were analyzed in 36 patients in total (in 2 subjects no tumor tissue have been found, 1 patient had cancer, but the information about the grade on the tumor was missing and 1 subject was source of adenoma, but not of tumor tissue). From the 32 donors left, 3 were categorized as G1 tumor, 14 as G2, other 14 as G3 and only one sample was classified as G4. Apart from them, for the rest of the investigated targets (genes), the data used for evaluation was derived from 29 patients. Two donors had no tumor tissue. From the remained 27 samples, 3 patients were diagnosed as G1 tumor, 13 as G2 and 10 as G3, see Table 3-4, page 18. Further, by using the data from each

grading-group, linear regression analysis was performed. No significant deviation from zero was observed when mRNA expression of the investigated genes was compared with the grade of the tumor (data not shown). In addition to the linear regression analysis, Kruskal-Wallis analysis was performed, to test whether the mRNA expression variations among the different grade-groups are significant. No significance was found (data not shown).

Nevertheless, a tendency of increased *OPN* expression in patients with moderately (G2) and poorly differentiated (G3) tumors compared to those with well differentiated tumors (G1) was detected (Figure 4-9). This observation was in contradiction with the reported significantly lower *OPN* expression in poorly differentiated (G3) tumors compared to those with well (G1) or moderately (G2) differentiated carcinomas, which was explained by considering that in highly undifferentiated tumors the loss of cellular differentiation could be responsible for an increased reduction of the *OPN* expression (129). However, to draw statistically safe conclusions, more patient's data per group must be collected and analyzed.

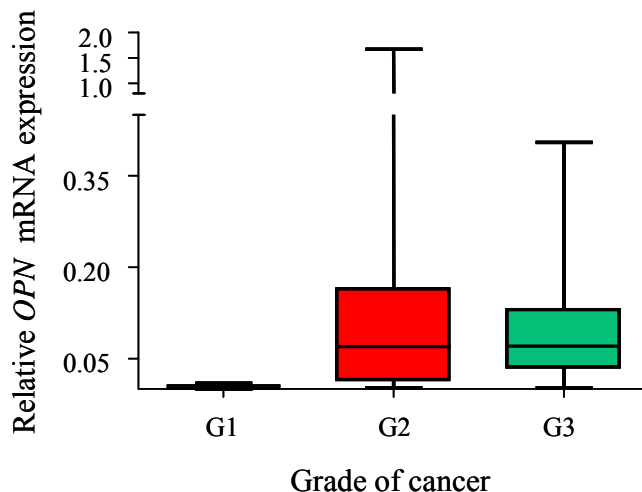


Figure 4-9 Tendency of increased *OPN* mRNA levels in G2 compared to G1. The graph shows the normalized level of *OPN* mRNA in samples from different tumor grades. The application of Spearman analysis, using each column of data, resulted in although not significant, but mild positive correlation coefficient of $r=0.3$, $p=0.14$. A $r=1$ gives a perfect correlation between the both data patterns analyzed.

4.3.4. Analysis of expression levels in carcinomas according to the tumor staging

The staging of the tumor was based on TNM system, where T specifies the extent of the tumor, N the extent of spread to the lymph nodes, and M shows the presence of metastasis. This TNM combination correspond to one of 5 stages, as stage 0 (S0) is carcinoma *in situ* (early cancer that is present only in the layer of cells in which it began) and stage 4 (S4) indicates that the cancer has spread to another organ.

In the present study, the expression of *DEFA 1-3* and *DEFA 6* in tumor have been analyzed in 36 donors, but the expression of only 32 subjects were tested for stage-dependency (see page 52). 8 of them had S1 tumor, 7 were with S2, 11 with S3 and 6 with S4 pathologically staged tumors. For the rest of the investigated genes, only 26 patients with defined stage of cancer were tested, as 8 of them had S1 tumor, 6-S2, 9-S3 and only 3 presented cancer with stage 4 (S4). The lack of more data makes the indication of some tendency in the expression

difficult. However, besides the Kruskal-Wallis test, linear regression analysis was performed. As the applied tests revealed, none of the genes investigated has shown any association between its expression and the stage of the cancer (data not shown).

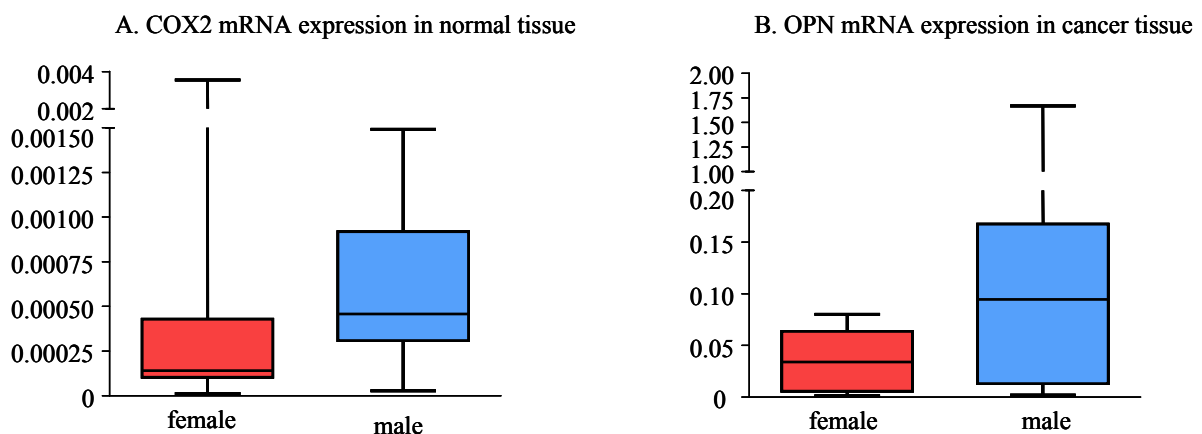
4.3.5. Gender and gene expression patterns in colorectal cancer

All collected data in this study was divided into groups of males and females. Thus the mRNA expression differences between the two gender groups in normal as well as in tumor tissue obtained from patients who underwent colonoscopy was investigated by using non-parametric Mann-Whitney test. The analysis revealed that the median expression of *COX-2* gene in normal tissue is significantly higher in males compared to females. Similar effect was observed regarding the median expression levels of *OPN* in tumor colon tissue (Table 4-5 and Figure 4-10). For the rest of the genes, no specific tendency in the median expression between male and females was observed (data not shown).

Gene	Mann-Whitney test, p-value	
	Normal tissue	Tumor Tissue
OPN	0,43	0,05
COX-2	0,024	0,51
DDX6	0,22	0,8
DEFA 1-3	0,8	0,21
DEFA 6	0,7	0,72
eIF4E	0,3	0,2
HAT1	0,14	0,5
HSP90B	0,23	0,28
PKM2	0,7	0,18
PLK1	0,46	0,35
USP28	0,74	0,51

Table 4-5 Comparison of the mRNA expression levels in males and females. mRNA expression levels of 12 GOI were analyzed. Mann-Whitney test was determining whether the expression differences between the both gender- groups were significant. Highlighted are the significant results.

Figure 4-10 Gender-dependent differences in the expression pattern of (A.) *COX-2* mRNA in normal colon tissue and (B.) *OPN* mRNA in tumor colon tissue. The black horizontal lines in the boxes represent the median expression in the different gender groups. The outcome of the statistical analysis is presented in Table 4-5



4.4. Interplay of 12 target genes in individual patients. “Bar code” representation

In the previous chapter the fluctuation in the expression profile of single target genes across all patients analyzed was investigated and it could be shown that at least some genes might be used to distinguish normal from tumor tissue. The predictive power was, however, not good enough for use in clinical applications. Therefore, in the present chapter, an attempt is made to use all 12 genes simultaneously for distinguishing normal from tumor tissues. Such genetic discrimination might support histological approaches as they are performed in the clinical routine. To facilitate such an analysis, for every individual donor, the combination of the expression patterns for the entire set of targets was visualized as a “bar code”.

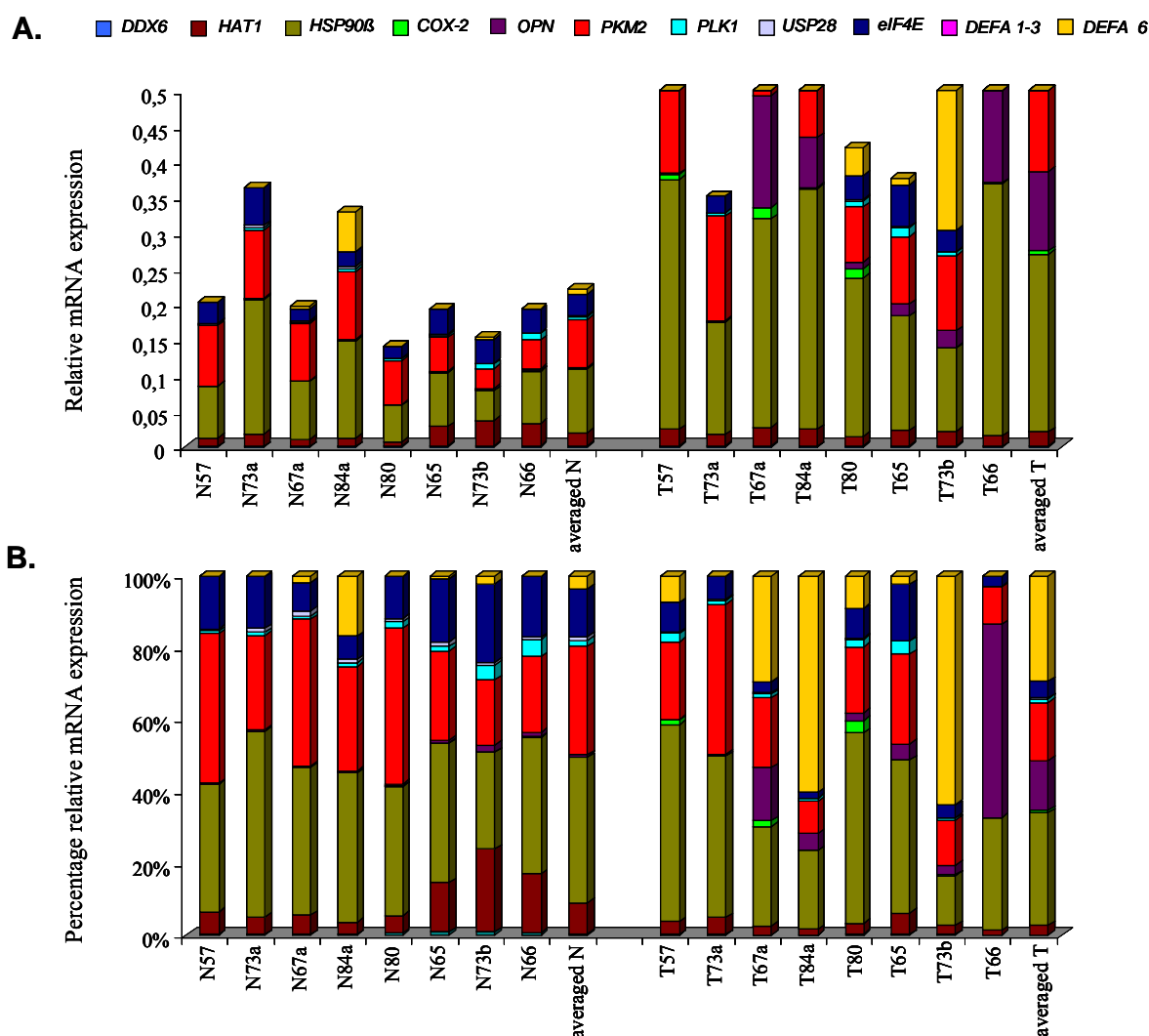
In order to present the meaning of the term “bar code”, the expression of 12 selected genes in 16 paired normal and tumor tissue samples, derived from randomly selected 8 patients, was evaluated. The outcome of the analysis is shown in Figure 4-11, page 55. Each column gives the sum of the expression of 12 genes in one individual donor for normal and tumor tissue, respectively. Each slice of the columns represents the contribution of a given single gene. Furthermore, the last column from each group represents the averaged expression among the 8 analyzed donors.

The analysis shows that in normal tissue, the height of the columns, respectively the single gene’s contribution in this combination of genes was almost similar (Figure 4-11, A). This observation reflects a distinct similarity in the expression pattern of the 12 genes in normal colon tissues. Figure 4-11, B. shows the same data as Figure 4-11, A., but presented as relative value of each gene in percentage. As in Figure 4-11 A., the values in normal tissue were rather similar. In contrast, the tumor colon tissues showed extensive variability in the expression patterns. Apart from the stable slight up-regulation of *HSP90 β* and *PKM2* genes, the expression variability appeared particularly due to the expression of *OPN* and *DEFA 6* and to a lesser extent, to the role of *COX-2* which essentially did not contribute to the columns of the respective normal tissues, but it is clearly seen in a relatively moderate percentage of the tumor tissues analyzed (Figure 4-11, B).

The expression variability in tumor is not a surprising observation since it is well known that neoplasms are predominantly heterogeneous and contain a variety of subpopulations of different cell types, with different metastatic potential (130). In addition to that, a recent study reported that for the case of colon cancer this inter-tumor heterogeneity is bigger than the intra-tumor heterogeneity (131).

Indeed, in an empirical approach (using the expression data derived from randomly selected 8 individuals, presented in Figure 4-11), by trial and error, the genes *PKM2*, *HSP90 β* , *OPN* and *COX-2* were found to be those with the major impact for the discrimination between non-malignant and malignant tissues. Later on it came out that this result was verified by the performed in addition Principal Component Analysis (PCA), (section 4.6, page 61), where the same set of genes was found to be the “driving force” for achieving discrimination between the non-malignant, “N” and malignant, “T” colon tissues and/or vice versa.

Figure 4-11 (A.) Cumulated gene expression of 12 targets in paired normal (N) and Tumor (T) tissue samples, derived from 8 subjects. The left group of columns shows the gene expression in normal tissue (N). The right group of columns shows the same set of data for tumor tissue (T). Every single patient is denoted with capital letter corresponding to the tissue type analyzed and number, designating the age of the investigated patient. In addition, the small types (letters) denote different patients with the same age. The last columns from each separated group illustrate the averaged expression across the subjects in the defined group. In contrast to the similar expression patterns in normal tissue, the columns showing the expression profiles of tumor tissues are markedly different in composition and in height; (B.) Percentage of contribution of the genes whose absolute expressions are shown in (A.).



Concluding remarks:

The presentation of the gene expression data as a “bar code” can be a useful strategy to perform preliminary selection of gene targets, which can take a part in a further computational analysis in order to test whether efficient discrimination between blinded paired normal and tumor colon tissue can be performed.

4.4.1. Discrimination of normal and cancer tissue in 18 patients with 78% accuracy and no false negatives

In order to verify the prominent role of *PKM2*, *HSP90β*, *OPN* and *COX-2* genes for their ability to distinguish non-malignant from malignant tissue, an independent set of 36 blinded samples, corresponding to paired normal and tumor tissues, derived from 18 patients was tested.

To carry out the blinded study, a separate criterion (E_{gC}), corresponding to any of the selected genes was designed. The E_{gC} for certain gene (g) was computed by summing up its average expression (\bar{E}), obtained from the normal tissue of the initial (learning) set of 8 patients and the corresponding standard deviation (SD). The combined gene expression data for the learning set of 8 patients was already presented in Figure 4-11, page 55.

$$E_{gC} = \bar{E} + |SD| \quad (1)$$

where \bar{E} and SD are calculated by equations 2 and 3 as follows

$$\bar{E} = \frac{1}{N} \sum_{i=1}^N E_i \quad (2)$$

$$SD = \sqrt{\frac{1}{N} \sum_{i=1}^N (E_i - \bar{E})^2} \quad (3)$$

where E_i is the relative expression of selected gene in sample i and N corresponds to the number of patients used for computation of the criterion. In our calculations N is equal to 8. The above performed calculation steps were applied to compute criteria for all four selected genes.

Next, the ratio between the relative expression (E_g) of the selected genes for defined, unknown tissue sample and the computed criterion (E_{gC}) of the corresponding gene (g) was calculated, see equation 4

$$R_g = E_g / E_{gC} \quad (4)$$

where R_g is the ratio generated for a selected gene in single/defined tissue sample.

In order to have comparable values, the estimated gene expression ratio (R_g) was converted to the fold-change (F_g) by using reciprocal transformation. Thus,

if $R_g \geq 1$, then $F_g = R_g$, while

if $R_g < 1$, then $F_g = 1 / R_g$

By this way, F_g adopts a positive value, when the expression of the selected gene in the defined tissue sample is higher compared to the corresponding criterion. Since the criterions

are based on the expressions of the genes in normal tissue, it is assumed that each positive value corresponds to a sample which revealed elevated gene expressions compared to that in normal tissue. This sample can be classified as a tumor. Correspondingly, F_g adopts a negative value, when the respective gene has a lower expression compared to the criterion. This would reflect the situation in the normal tissue.

Next, in order to summarize the total expression impact of all selected genes in one unknown sample, the corresponding reciprocal values were summed up.

$$\text{SUM} = F_{PKM2} + F_{HSP90\beta} + F_{OPN} + F_{COX-2} \quad (5)$$

where the index of each F indicates the gene for which this value was created.

According to equation 5, a sample with a positive sum will be taken as cancerous (malignant) tissue, and respectively, sample with negative sum will be regarded as normal tissue. This procedure was applied to 36 blinded samples, derived from 18 patients. The outcome from the computational analysis is shown in Table 4-6, page 58. It demonstrated that 28 out of 36 samples, i.e 77.78% were correctly predicted with no reservation regarding the final outcome, and 8 samples, i.e 22.22% were incorrectly predicted (Table 4-6), marked in grey). Interestingly, they were all false positive, which means that they were estimated as tumor, although originally were derived from normal tissue. This fact can be counted as a distinct advantage of the analysis. It is for the sake of the false positive patient to undergo additional examinations and thus indeed to be verified that the examined tissue is normal and not tumor than to predict tumor as a normal tissue and skip the further examination procedures. On the other hand, in this analysis, not a single tumor would have escaped detection.

Table 4-6 Listed outcomes from blinded study where 36 randomly selected samples from paired normal (N) and tumor (T) colon tissues were predicted according the defined criterions. The computed fold changes for the set of the selected genes (Fg), the sum of them and the predicted answers are listed. The latter were compared with the correct answers. Blinded sample, which possesses positive value for the sum of Fg is charged as a tumor and the sum with negative value, regard the sample as a normal. The marked in grey cells depict the wrong predicted blinded samples (22.22%). 77.77% were correctly predicted.

Sample	F _{HSP90β}	F _{COX-2}	F _{OPN}	F _{PKM2}	SUM	Predicted Answers	Actual Answer
1	2.88	9.69	23.95	1.56	38.10	T	T
2	1.31	1.33	-1.02	1.29	2.90	T	T
3	1.72	-1.68	2.32	1.13	3.49	T	N
4	1.03	-1.96	-5.25	1.06	-5.12	N	N
5	-1.69	1.79	-5.92	-1.78	-7.60	N	N
6	-1.83	3.59	30.35	-2.68	29.43	T	T
7	1.78	4.43	65.75	1.38	73.34	T	T
8	1.45	7.82	2.23	-1.11	10.39	T	N
9	-1.67	-1.06	-1.24	-2.09	-6.06	N	N
10	1.99	29.50	45.06	1.16	77.72	T	T
11	1.71	-1.53	1.49	1.79	3.46	T	T
12	1.88	1.96	37.24	1.22	42.30	T	T
13	2.06	11.06	176.5	1.82	191.5	T	T
14	-1.40	-16.4	-34.7	-1.72	-54.2	N	N
15	-1.60	-37.1	-3.33	-1.95	-44	N	N
16	2.84	7.88	34.98	1.59	47.29	T	T
17	-2.21	2.04	3.48	-1.05	2.25	T	T
18	1.08	-1.27	1.09	-1.39	-0.48	N	N
19	-2.11	-1.28	1.67	-1.84	-3.56	N	N
20	1.16	1.23	4.07	-1.03	5.42	T	N
21	1.18	9.45	25.09	-2.15	33.56	T	T
22	-1.09	10.43	31.10	-1.48	38.95	T	T
23	2.42	2.39	19.26	1.21	25.28	T	T
24	2.56	8.04	77.65	2.41	90.65	T	T
25	1.53	-3.86	1.50	-1.06	-1.88	N	N
26	2.07	2.07	2.36	1.41	7.91	T	N
27	-1.27	-1.88	8.63	-1.93	3.53	T	T
28	1.73	-1.05	1.49	1.14	3.31	T	N
29	1.18	1.08	2.41	-1.06	3.59	T	N
30	2.46	-7.71	-1.01	2.98	-3.26	N	N
31	-1.54	2.31	-1.58	-2.74	-3.50	N	N
32	1.19	1.32	1.31	-1.48	2.33	T	N
33	2.13	3.65	3.31	1.04	10.14	T	T
34	3.37	31.62	728.5	1.67	765.2	T	T
35	7.26	1.72	50.69	1.64	61.31	T	T
36	-1.19	1.98	3.25	1.03	5.09	T	N

Concluding remarks:

By trial and error, four genes were found to have a major impact for discrimination between non-malignant and malignant tissues, namely PKM2, HSP90β, OPN and COX-2. Based of their expression patterns, an empirical approach has been developed. The latter accomplished correct prediction for the type of tested tissue of about 78%.

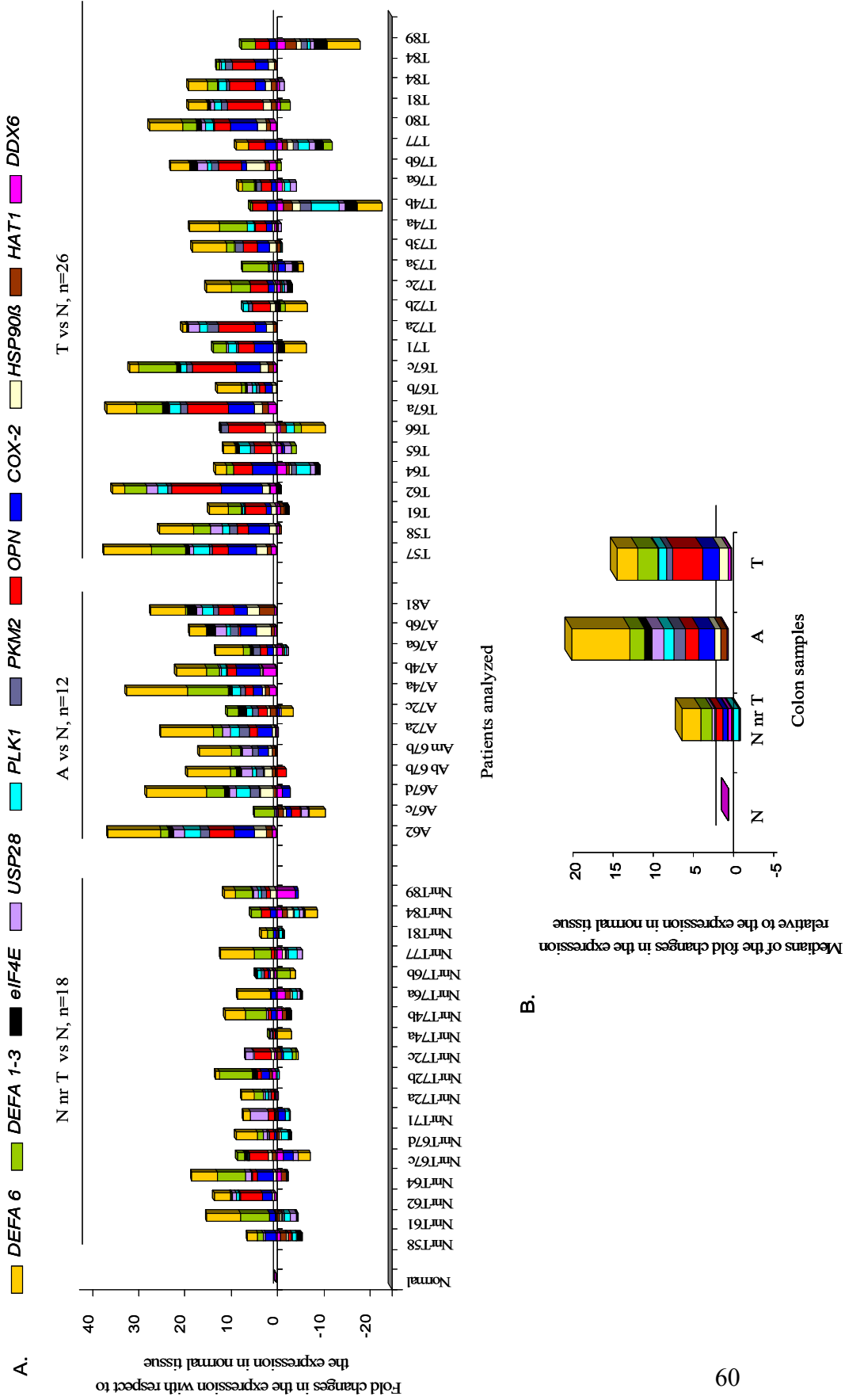
4.5. Combined gene expression analysis of 12 GOI in Normal, Normal near Tumor, Adenoma and Tumor colon tissue across 26 analyzed donors. Application of the Relative Expression Software Tool (REST)

By applying REST, group-wise comparison of the relative expression of 12 selected genes in 4 different types of colon tissue obtained from 26 donors in total was performed. Non-malignant, normal (N) tissues and malignant, tumor (T) colon tissues were derived from all 26 donors, with histologically confirmed stages of colon cancer development. 18 of those objects provided normal near tumor (NnrT) and 12 of them - adenoma (A) tissue samples.

For more detailed information about how the software works, refer to 3.2.5.3, page 31. In brief, after gene expression normalization, the REST software determines the expression alterations in the “sample groups” compared to the respective expressions in the “control group”. As a consequence, the evaluated data for each individual donor was plotted as a “bar code” presentation. For better display, the data was log transformed with base 2. (Figure 4-12, page 60). In this graph, each colored slice of the column represents the ratio between the expressions of certain gene in the “sample group” and the respective expression in the “control group”. The first column of data, plotted on x-axis, symbolizes the expression of the entire gene set in the normal tissue, the so-called “control group”. In fact, the expression in the “control group” is equal to one since it was taken as a baseline in order to compute the expression’s changes in the “sample groups”. Therefore, when is log-transformed the expression in the normal tissue crosses the zero line. Further, in Figure 4-12, page 60 the “bar code” columns for individual donors with “NnrT”, “A” and “T” tissues are consecutively shown. The distinct tissue’s groups are separated from each other by empty space. In addition, for simplicity and better overview Figure 4-12 B, page 60 summarized the medians of the log-transformed expression ratio computed across the donors for each defined tissue’s group. The slices of the bars above the zero line correspond to the genes which demonstrated elevated expression, compared to the expression in the “control group”, namely the normal tissue. Conversely, the genes exhibiting down-regulation in comparison to the expression in normal “N” tissue are presented as slices below the zero line. The slices close to the zero represent genes with no change in the expression.

The outcome from the REST computation revealed that the “NnrT” tissues, surgically removed from location approximately 5 mm away from the tumor and which are macroscopically proven to be benign showed already alterations in their expression profiles compared to the “N” colon tissue, removed at a distance 20 to 50 cm from the tumor side. This is in consistence with several previous studies where the presence of genetic changes in morphologically normal tissues derived from cancer patients was already reported. Loss of heterozygosity or microsatellite instability were some of the observed alterations. However, it is still unclear whether these genomic modifications represent early precursors of the cancer, markers of increased risk, or simply population based polymorphisms (132; 133).

Figure 4-12 (A.) Comparison of the combined expression patterns of 12 GOI across all 26 investigated patients. Displayed are the expression ratio between NnrT and N, A and N and T over these patients. These ratios represent the gene expression changes respectively in NnrT, A and T tissues when compared to the expression in N tissue. The data is log-transformed with base 2. The first single column represents the expression in N tissues, the log of which is zero. The different ratio-groups are separated from each other by empty space. Single slice in an individual column presents the expression changes of a given gene, compared to expression in N tissue, and its contribution in the combination of selected target genes; (B.) Summarized presentation of the medians across the log-transformed expression changes in NnrT, A and T colon tissues versus the expression in N tissue; n indicates the number of analyzed tissue samples.



Nevertheless, the presented preliminary data is very useful, since the careful characterization and examination of the observed genetic alterations is necessary to understand the molecular changes possibly leading up to colon cancer development.

The analysis showed that apart from the substantial inter-individual variations in the expression patterns, relative elevation in the expression of *DEFA 1-3*, *DEFA 6* and to some extent *OPN* and *COX-2* genes can be detected in “NnrT” tissues. These targets remained up-regulated also in “A” tissues (Figure 4-12, A and B, page 60), where the expression of *DEFA 6* seems to be the hallmark for this tissue type (see detailed description in section 4.3, page 43).

Notable is that all analyzed “A” samples were benign, although over time they can become malignant. Due to that reason, the detailed examination of their expression patterns would be an additional, useful source for better understanding the biology of the cancer.

Furthermore, the expression of *DEFA 1-3*, *DEFA 6*, *OPN* and *COX-2* genes remained elevated also in “T” tissue. Particularly over all patients, *OPN* and *DEFA 1-3* expressions were higher in the “T” when compared to the respective expression levels in the “A” tissue (Figure 4-12, B, page 60). The elevated expression of these 4 genes during the cancer progression is not surprising, since many reports already proved their specific role in promoting the cell migration and proliferation (29; 34; 40; 97). Those features are typical for any malignant cell or cells derived from fast growing tissue as the adenoma.

In addition to the up-regulation of *DEFA 1-3*, *DEFA 6*, *OPN* and *COX-2*, the expression profile of the examined 12 adenoma samples showed, relatively low but stable increase in the expression of *PKM2*, *PLK1* and *HSP90 β* , which stayed relatively unchanged in cancer, “T” colon tissue. This observation is expectable since some of the investigated genes take part in processes such as cell proliferation, like *PLK1*, which drives the cells into mitosis and/or *PKM2*, which is target involved in glycolysis or as *HSP90 β* molecule, which is maintaining the homeostasis of the malignant cells in the hostile environment. All of those processes support the development of cancer.

Concluding remarks

REST computational analysis, combined with “bar code” representation is promising tool for presentation and evaluation of the expression data.

4.6. Tissue discrimination by using Principal Component Analysis (PCA).

In addition to the empirical “bar code” representation, simultaneous examination of the selected genes was performed by using PCA, (for detailed description refers to 3.2.5.2, page 25). This statistical technique provides a way to achieve one of the main goals of this study, namely to identify those expression patterns, which most efficiently discriminate non malignant tissue from tissue with a defined stage of cancer. As a result, PCA allows a safe identification of the genes with the highest diagnostic value. The analysis was carried out using the raw, not normalized expression data from all analyzed genes (including also the

reference genes) across all types of investigated tissues (normal, normal near tumor, adenoma and cancer). The outcome computed from the PCA is displayed in Figure 4-13, page 63. The different colors correspond to different types of tissue. The two dimensional image represents the first 2 out of 13 Principal Components (PCs) obtained after implementing the assay. In principal, the number of the computed PCs is equal to the number of variables (genes) analyzed. In the present study, however, the number of the PCs did not reflect the number of analyzed genes. The reason for this discrepancy was the mentioned above fact that only one primer pair was used to assay the expression of 2 different genes with almost identical nucleotide sequences (DEFA 1 and DEFA 3). As a consequence, one of the 13 variables accounted for two investigated genes (DEFA 1 and DEFA 3). Thus, PCA computed 13 PCs corresponding to 14 genes (12 GOI and 2 HKG). Nonetheless, among the entire set of PCs, only the first two PCs are the most important one, possessing the highest eigenvalues. The eigenvalue measures the variance in all the variables which is accounted for by that component. The higher the eigenvalue, the larger the importance of the corresponding principal component should be. Thus, in Table 4-7 are listed the variance percentage and the respective eigenvalue for the entire set of principal components computed.

PC	Eigenvalue	% variance
1	29.01	50.05
2	13.95	24.06
3	7.1	12.24
4	3.39	5.84
5	1.60	2.76
6	0.94	1.63
7	0.73	1.26
8	0.41	0.71
9	0.28	0.49
10	0.24	0.42
11	0.15	0.26
12	0.09	0.15
13	0.07	1.12

Table 4-7 List of the eigenvalues and the corresponding percentage of total variance accounted for by each principal component (PC) computed from the PCA of 14* analyzed genes across normal, normal near tumor, adenoma and tumor colon tissue.

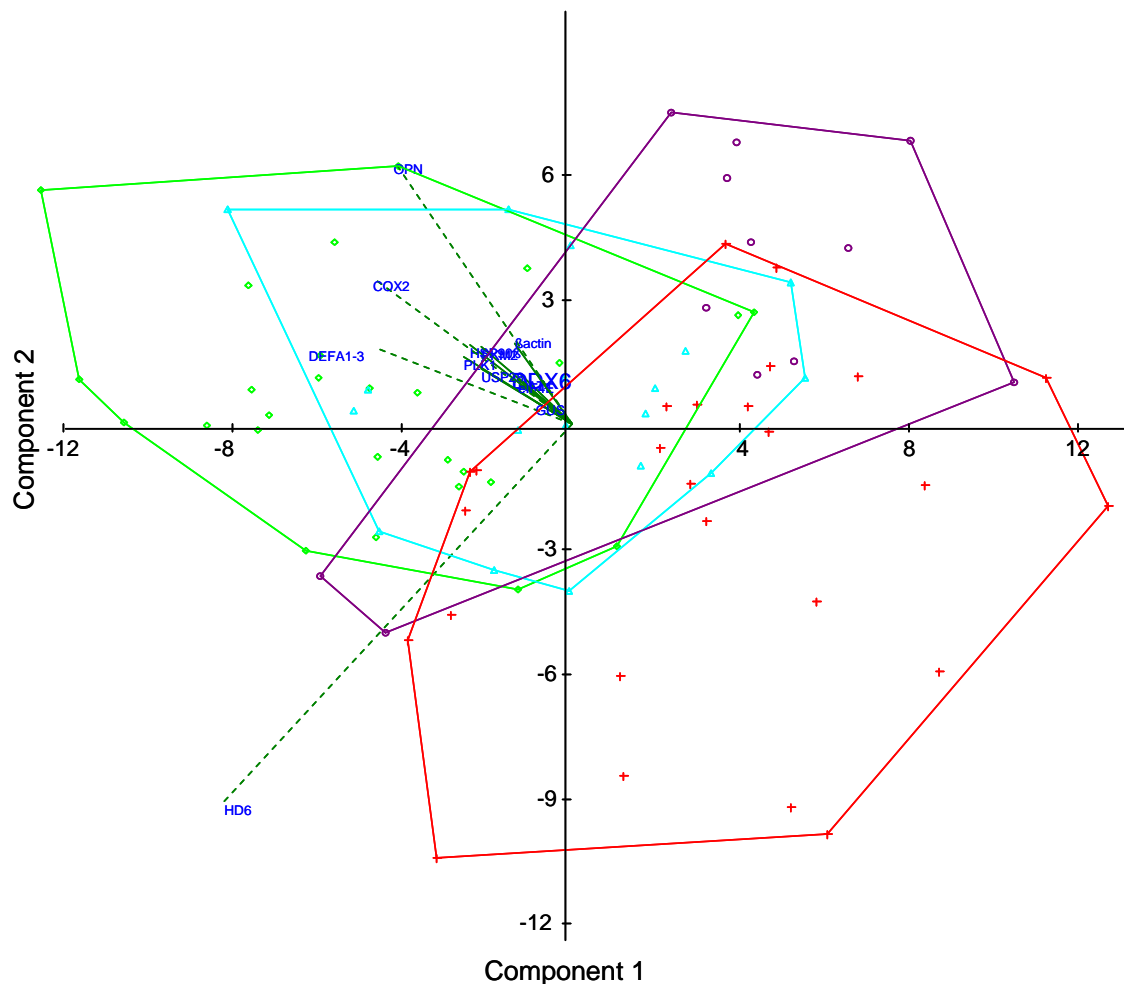
The principal component is a suitable combination of all 14 investigated genes with different weights (In Figure 4-14, page 64, it is shown how the 14* investigated genes contribute to PC1 and PC2, the most important principal components).

* Notice that one primer pair was designed for simultaneous detection of 2 genes (DEFA 1 and DEFA 3), because of their almost identical nucleotide sequences.

The first two principal components accounted for 74% of the total variance in the original data and therefore they had the greatest impact in finding differences among the populations of the four different tissue types.

In addition, an analysis of the loading of the two most informative principal components was performed to test, as to what degree the different original variables (genes) contributed to these components. The analysis showed that *DEFA 6*, *DEFA 1-3*; *COX-2*, *OPN* and to a lesser extend *HSP90 β* and *PLK1* were the ones with the highest predictive contribution for tissue discrimination. The rest of the genes can be regarded as less critical for the further analysis (Figure 4-14, page 64).

Figure 4-13. The Scatter Plot of the first Principal Component (PC1) versus the second most important Principal Component (PC2). Evaluated was the data from four different types of tissue. The green represents the “normal”, non-malignant population; blue – the “normal near tumor”; violet – the “adenoma” and the red color - the “tumor”, malignant group of colon tissues. A teaching version of this figure is given in the Material and Methods section, 3.2.5.2, page 25.



The outcome of the PCA revealed a clear tendency of separation between the “normal”/“non malignant” colon tissue population (colored in green) and the respective “tumor”/“malignant” tissue population (colored in red) since the overlapping region between these two populations was relatively small. In addition, the trend of the adenoma population (colored in violet) to move in direction similar to that of the “tumor” one was also observed. Nevertheless, reasonable/solid conclusions about the discrimination of the “normal near tumor” (colored in light blue) and “adenoma” (colored in violet) populations either from “normal” or from “tumor” populations was not worth to be built up. Therefore, these tissue populations were excluded from additional examinations.

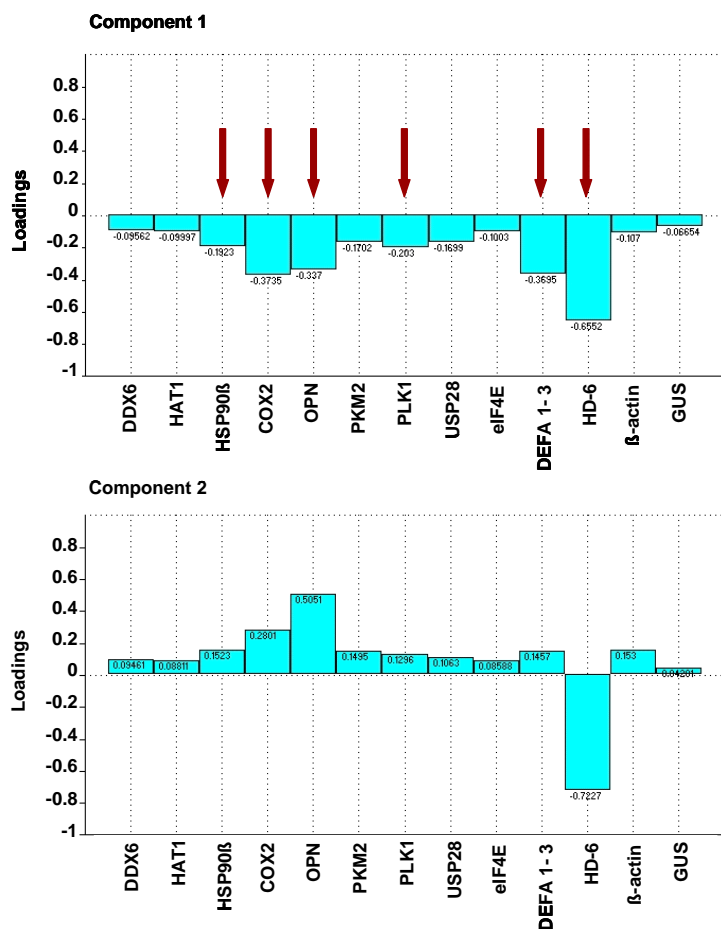


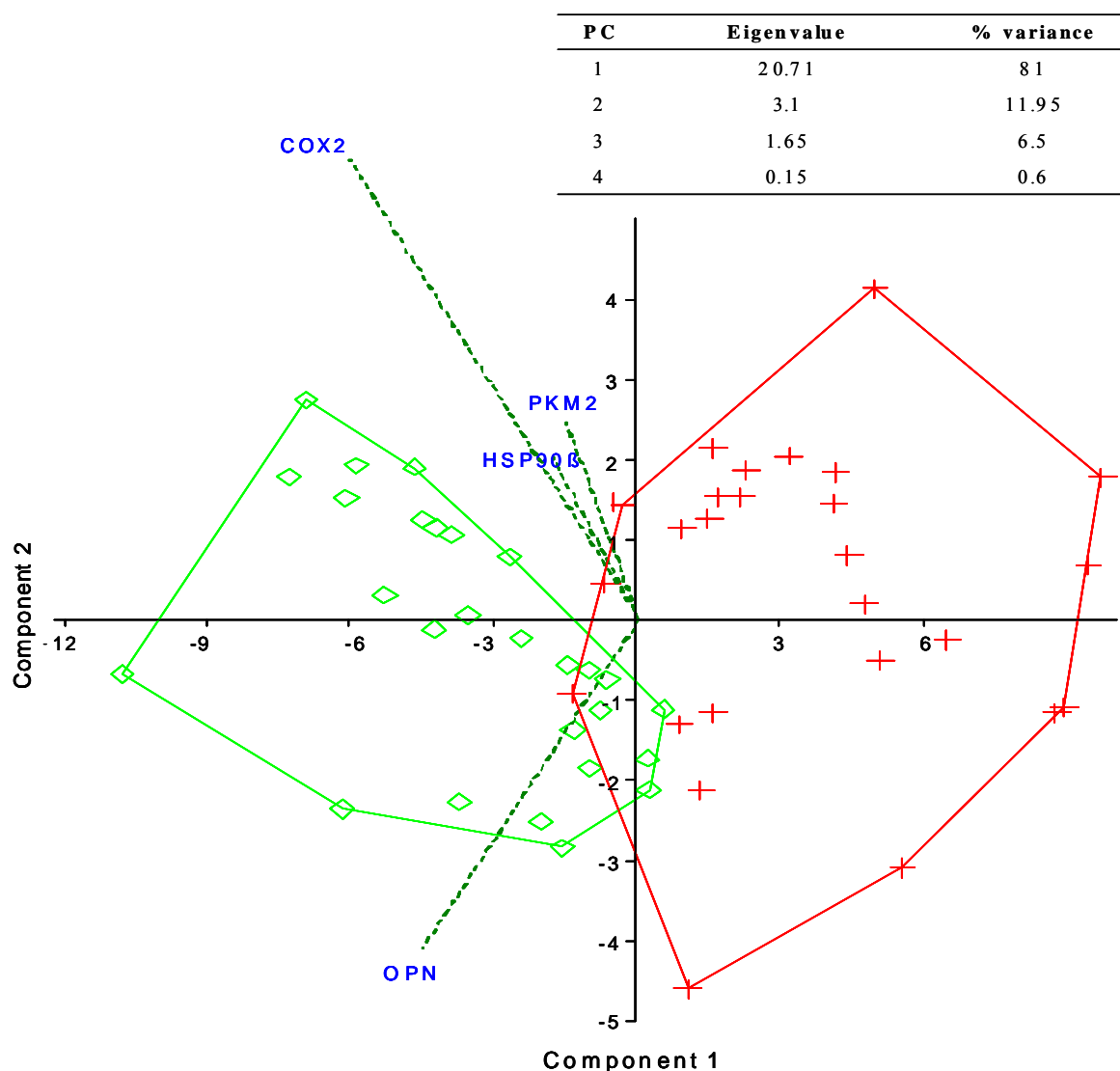
Figure 4-14 Combination of the 14* examined genes to PC1 and PC2, i.e. loading profiles. The higher the loading of the variables, the more this variable contributes to the variation accounted for by the principal component. The red arrows indicate the genes with highest weight.

* *DEFA 1-3* corresponds to two genes

COX-2, OPN, HSP90β and PKM2 safely distinguished normal from tumor tissue!

The further efforts using PCA, focused on finding a set of genes by which the best discrimination between “normal/non-malignant” and “tumor/malignant” tissue population can be achieved. In order to do so, various sets of different combinations were assayed by PCA. Indeed, the best separation between those two populations was obtained when the expression of *COX-2*, *OPN*, *PKM2* and *HSP90β* were simultaneously analyzed. The PCA result is reported in Figure 4-15, page 65 together with the eigenvalues and variance percentages of the first two PCs which account for by 92% of the total variance of the original data.

Figure 4-15 Scatter Plot of the two most prominent principal components (PC). For performance of the computational analysis, gene expression data of 4 genes, namely *COX-2*, *OPN*, *PKM2* and *HSP90 β* across “normal” and “tumor” populations of colon tissue have been applied. The green block represents the “normal” population, the red – the “tumor” population. In addition, eigenvalues and the corresponding percentage of total variance accounted for by the first two most prominent PC are listed.



One can immediately recognize that even in this two dimensional representation the large majority of data points is either in the red field (for cancer tissue) or the green field (for normal tissue). In a multidimensional representation, the discrimination power is even better. This result validates previous observations in this study, where by applying empirical approach, with trial and error, completely the same (identical) set of genes was selected as the most prominent one, achieving the best discrimination between malignant and non-malignant colon tissue 4.4.1, page 56.

Concluding remarks:

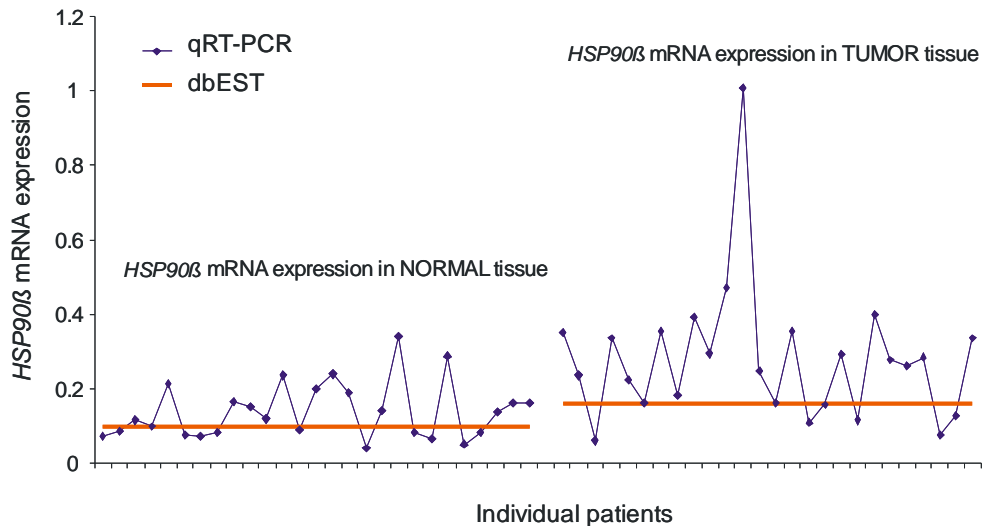
PCA was applied as a tool to analyze the multivariate gene expression data derived from normal, normal near tumor, adenoma and tumor colon tissues. The analysis verified that the selected combination of genes, namely HSP90 β , OPN, COX-2 and PKM2, is sufficient to significantly distinguish “normal/non-malignant” from “tumor/malignant” colon tissue populations. This finding is in consistence with an observation, previously reported in the present study, where empirical approach, with trial and error, achieved the same outcome.

4.7. Direct comparison of gene expression data generated by qRT-PCR with the data derived from NIH’s data base dbEST.

The present investigation aims to test the predictive power of gene expression data derived from NIH’s data base dbEST, which collects expression results from a large number and variety of DNA microarray experiments. Therefore, a comparative analysis was performed between the gene expression data generated by qRT-PCR, across 26 paired/matched normal and tumor colon tissues, and the data obtained from the dbEST data base. Presently, dbEST provides information only for normal and tumor colon tissue. Unfortunately, data from intermediate tissues, such as normal near tumor and adenoma is not available and therefore they were not included into the analysis. Under investigation was the entire set of analyzed genes, except *DEFA 1-3* and *DEFA 6*. No information about their expressions was found in the data base, neither for normal nor for tumor colon tissues.

In order to carry out the comparative analysis, normalization of the data had to be implemented. According to the observations earlier discussed in this work (see section 4.1, page 35), β -actin was used for calibration of the data obtained either by dbEST data or derived by qRT-PCR. The quantification of the relative expression was based on the expression ratio of the target gene versus the selected reference gene. Once the matter of normalization was clarified, the data for other genes can be taken without further precautions, except that they have to be obtained with the identical procedure as the β -actin values. Therefore, a direct comparison of patient data, generated by qRT-PCR with the data base values could be performed. Generally, one can expect that normalized expression values for different genes are vastly diverse among patients. This is shown in Figure 4-16, page 67, where as an example, the variety in the *HSP90 β* expression pattern (derived by qRT-PCR) across 26 patients is presented in combination with the averaged expression obtained from dbEST data base.

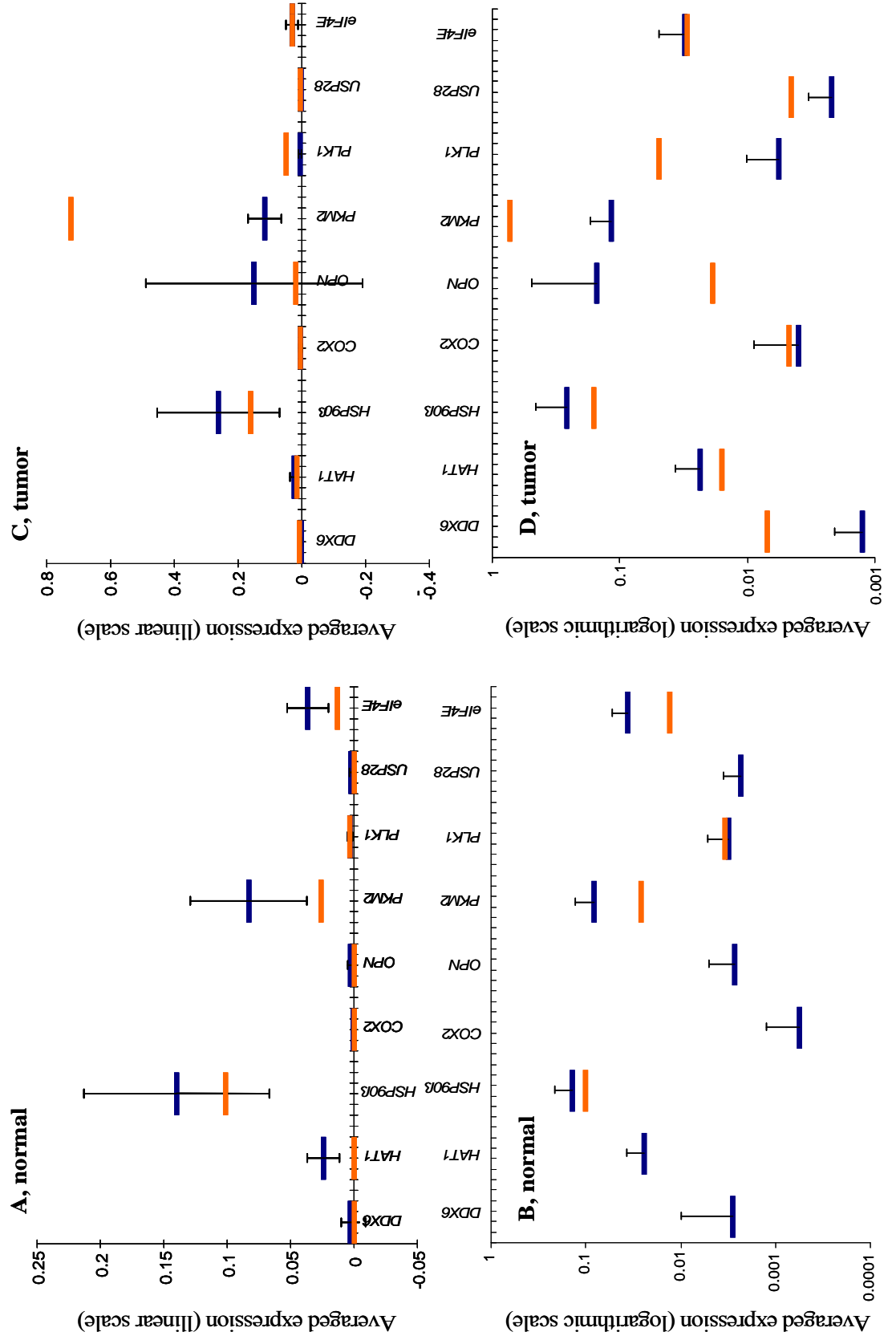
Figure 4-16 *HSP90β* expression data generated by qRT-PCR and the averaged data provided from dbEST data base in normal and tumor tissue. Every rhomb symbol (◆) depicts the *HSP90β* expression in one individual, obtained by qRT-PCR. For better visualization, the rhombs are connected. The left part of the graph illustrates the expression in normal tissue, the right - in tumor. The orange horizontal line represents the average values derived from dbEST database for hundreds to thousands of patients.



Apart from the inter-individual variability and in order to facilitate the comparison between the gene expressions derived from the qRT-PCR and the dbEST data base, averaged gene expressions from the both different sources were combined and plotted in Figure 4-17, page 68. In addition to the presentation of the data on a linear scale, for better visualization, the data was introduced on a logarithmic scale as well. This additional way for data description was necessary since, in fact the gene expression levels provided by qRT-PCR and dbEST data base, covered differences in the expression by more than 3 orders of magnitude. Therefore, by log transformation this wide range of values can be reduced and presented in a more manageable range.

So far, the analysis revealed that dbEST database had no entries for the expression of *DDX6*, *HAT1*, *USP28*, *OPN* and *COX-2* genes in normal tissue, while such data were available in regard to tumor colon tissue. Therefore, their expressions were counted as a zero. Thus, in Figure 4-17 A., page 68 the presented expression of these genes appeared on the zero line, while when the data was presented on logarithmic scale (Figure 4-17, B) these genes were not depicted at all. One plausible reason for the missing data in the database would be their relatively low transcriptional levels compared to the high expression level of well abundant genes. Therefore, their detection was impossible in terms of the applied DNA microarray approach, the source of the data in dbEST database. Such a low abundance (of the above mentioned genes) in normal tissue is in consensus with previous studies where, in the case of *OPN* and *COX-2*, the hardly measurable expression in normal tissue can be rapidly induced by tumor promoters, oncogenes, growth factors, cytokines and other stimuli (134-136).

Figure 4-17 Comparison between the expression data generated by qRT-PCR and the data provided from dbEST data base. The averaged qRT-PCR gene expression data of all patients plus the standard deviations of each mean are shown as a smooth blue line. The smooth orange lines represent the average values derived from the dbEST database for hundreds to thousands of patients. (A) and (B), display the expression of all selected genes in normal tissue using linear and logarithmic scale, respectively. The corresponding expression in tumor tissue is plotted in (C) (linear scale) and (D) (logarithmic scale), respectively.



Similarly, the expression levels for those genes in normal tissue, measured by qRT-PCR were relatively low, when compared to the rest of the analyzed genes. This is more observable when the data were plotted on a linear scale (Figure 4-17, A., page 68). Thus, indeed the agreement between dbEST and qRT-PCR for those genes in normal tissue was quite reasonable. Moreover, the agreement in the case of *HSP90 β* , *PKM2*, *PLK1* and *eIF4E* genes was excellent. The difference in the expression between the qRT-PCR and dbEST sources was relatively low (between 1 to 3 folds), which in term of the different approaches used for determination on the gene expression is already quite precise.

Further, the predictive power of dbEST for cancer tissues is somewhat lower, but still good. One explanation, which could elucidate those disagreements, is that the database averages huge data over all subclasses of colon cancer, while the qRT-PCR data were generated from a cohort with a narrower spectrum of colon cancer stages. The estimated differences in the determined expressions between the qRT-PCR and the dbEST data base for tumor, in respect to *PLK1*, *PKM2*, *OPN* and *DDX6* genes were from 5 to 8 fold. Nevertheless, these poor results are still useful, because although not considerable, for almost all genes the expression's tendency between normal and tumor colon tissue remains the same, no matter which source has been used for evaluation. However, in respect to *DDX6*, general discrepancy in the expression profile, between normal and tumor tissue, was observed. In contrast to dbEST, where *DDX6* was increased in malignant tissues, according to the qRT-PCR approach, *DDX6* showed to be down-regulated in cancer when compared to the normal tissue (among 26 donors). In spite of that fact, the agreement with the database for the rest of the analyzed genes (*HAT1*, *HSP90 β* , *COX-2*, *USP28* and *eIF4E*) was very good, as the prediction was quite precise, in a range of 2 fold differences.

4.8. Gene expression in cell lines

In vitro cell models play an important role in understanding cellular events related to (patho)physiological conditions in humans. Therefore, they are widely used to address specific research questions, though their cellular metabolism alters drastically compared to physiological conditions (137). Furthermore, the cultured cells are relatively simple working models in comparison with the *in vivo* system, where the influence of different neighboring cells, neuro(endocrine) regulators and blood flow is present. So far, it is not clear whether cells in culture retain the gene expression profiles of their *in vivo* counterparts. A study analyzing the gene expression patterns in 60 human cell lines, including colon cancer cell lines, disclosed that the tissue, the cell lines are derived from, is the main factor accounting for the variation in the gene expression (138). Moreover, a recent report designed to compare protein expression profiles of two colorectal cancer cell models with their *in vivo* counterparts revealed that despite the consistence in the expression of several biologically significant proteins, many discrepancies appeared (137). Thus, care should be taken to prevent misinterpretation on *in vitro* obtained findings when translating them to the *in vivo* situation.

In the present study, gene expression analysis of all analyzed targets was performed in three colon cancer cell lines, representing different stages of carcinogenesis, namely one colorectal adenoma (LT97), and two carcinoma (HT29, SW480) cell lines (Table 3-1, page 16). qRT-PCR was applied to three different extractions from the same cell line as the gene expression pattern of each extraction was measured in duplicate or triplicate. The expression of β -actin gene was used for normalization. The outcome of the analysis is shown in Figure 4-18. The fold changes in the expression among the lines as well as the estimated significance according to the performed One-way ANOVA test followed by Tukey post test are listed in Table 4-8, page 71.

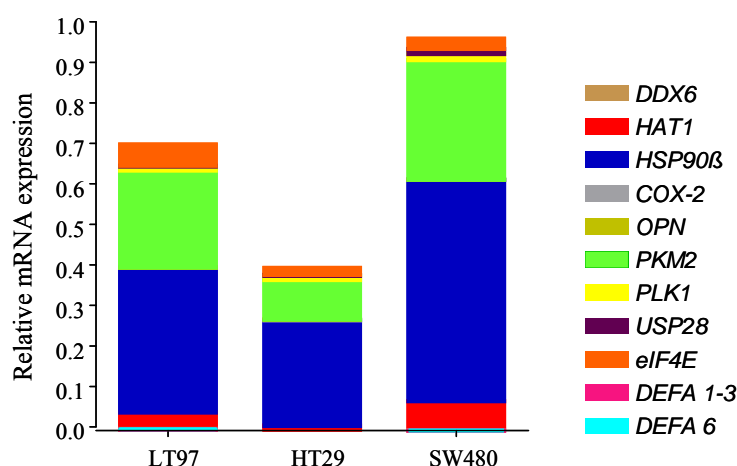


Figure 4-18 Cumulated averaged gene expression of 12 genes of interest in three cell lines, representing different stages of carcinogenesis.

According to the test, the highest significant expression differences were found in regard to *COX-2* gene. SW480 cell line showed extremely low *COX-2* expression compared to LT97 and HT29 cell lines with almost 39 fold and 57 fold, respectively. Furthermore, a significant moderate increase of 6 fold was observed for *USP28* and *HAT1* in SW480 compared to HT29. *USP28* was also 6 fold significantly elevated in SW480 in contrast to LT97. Another gene which showed extremely high expression alterations was *OPN*. Considerably high although not significant was the down regulation of *OPN* (with 66 fold) in SW480 compared to HT29. On the other hand, HT29 exhibited 36 fold (insignificant) increases in *OPN* expression versus LT97.

Apart from *DEFA 1-3* and *DEFA 6*, which are separately described below, the rest of the genes showed relatively small alterations in their expression, and most of them were insignificant (Table 4-8, page 71).

Table 4-8 Listed are the fold changes in the gene expression when two different cell lines were compared as well as the outcomes of the performed One-Way ANOVA test followed by Tukey post test. Marked in grey are the differences computed as significant; N/A- not available data

*for the case of *DEFA 1-3* only 2 groups had to be compared, therefore unpaired t-test was used

	Comparison between two different cell lines	Fold changes	One way Anova, with Tukey post test
<i>PLK1</i>	HT29 vs LT97	1.05	P > 0.05
	SW480 vs LT97	1.54	P < 0.05
	SW480 vs HT29	1.46	P > 0.05
<i>COX-2</i>	HT29 vs LT97	1.47	P > 0.05
	SW480 vs LT97	-38,6	P < 0.05
	SW480 vs HT29	-56,5	P < 0.01
<i>PKM2</i>	HT29 vs LT97	-2.44	P > 0.05
	SW480 vs LT97	1.23	P > 0.05
	SW480 vs HT29	3.00	P > 0.05
<i>OPN</i>	HT29 vs LT97	35.62	P > 0.05
	SW480 vs LT97	-1.86	P > 0.05
	SW480 vs HT29	-66.27	P > 0.05
<i>HSP90β</i>	HT29 vs LT97	-1.37	P > 0.05
	SW480 vs LT97	1.53	P < 0.01
	SW480 vs HT29	2.1	P < 0.001
<i>HAT1</i>	HT29 vs LT97	-3.15	P > 0.05
	SW480 vs LT97	2.1	P < 0.05
	SW480 vs HT29	6.55	P < 0.01
<i>USP28</i>	HT29 vs LT97	1.06	P > 0.05
	SW480 vs LT97	6.02	P < 0.001
	SW480 vs HT29	5.69	P < 0.001
<i>eIF4E</i>	HT29 vs LT97	-3,07373	P < 0.01
	SW480 vs LT97	-2,11274	P < 0.05
	SW480 vs HT29	1.45	P > 0.05
<i>DDX6</i>	HT29 vs LT97	-1,90	P > 0.05
	SW480 vs LT97	-1,39	P > 0.05
	SW480 vs HT29	1.37	P > 0.05
<i>DEFA 1-3*</i>		N/A in HCT116	
		N/A in LT97	
		N/A in HT29	
		N/A in SW620	
	SW480 vs Caco 2	-1.1	P > 0.05
<i>DEFA 6</i>		N/A	
		N/A	
		N/A in HCT116	
		N/A in HT29	
	SW480 vs LT97	-25.25801312	P > 0.05
	Caco 2 vs LT97	-191.0804709	P > 0.05
	SW620 vs LT97	-5222.838032	P > 0.05
	SW480 vs Caco 2	-7.565142593	P > 0.05
	SW620 vs SW480	-206.7794488	P > 0.05
	SW620 vs Caco 2	-27.33318589	P > 0.05

***DEFA 1-3* and *DEFA 6* expression in cell lines**

Since HT29 lacked the expression of all defensins analyzed (*DEFA 6* and *DEFA 1-3*) as well as *DEFA 1-3* was missing in LT97 cells, the expression of those targets was screened in three additional colon cancer cell lines, namely - Caco 2, HCT116 and SW620 (Table 3-1, page 16).

Besides the lack of *DEFA 1-3* in LT97 and HT29, the gene was also not detected in HCT116 and SW620 cell lines. In agreement with Zou et al. very low, but still measurable, was the *DEFA 1-3* transcription in SW480 (92). However, the expression was several orders of magnitude lower than the one detected in peripheral blood mononuclear cells (PBMCs). Only in one out of three Caco 2 extractions, *DEFA 1-3* was found. Furthermore, similar to the observations of several groups (97; 139), deviations in the levels of expression not only among the different cell lines, but also in extractions from different time periods within one cell line were detected. Still the reason for these alterations in the transcriptional regulation of *DEFA 1-3* is unknown. However, Agerberth et al. suggested that this could be explained by instability of an important transcription factor needed for alpha –defensin expression (139).

In regard to *DEFA 6*, additional to its lack in HT29, the gene was not found in HCT116 and PBMC. The latter observation was also not surprising, because the Paneth cells and not the leukocytes are the main source of *DEFA 6* expression (94). Among the rest of the analyzed cell lines, LT97 cell line, which possesses genetic alteration characteristics of early adenoma *in vivo* (Table 3-1, page 16), revealed higher *DEFA 6* expression. In SW480, Caco 2 and SW620 cell lines the defensin transcription was lower.

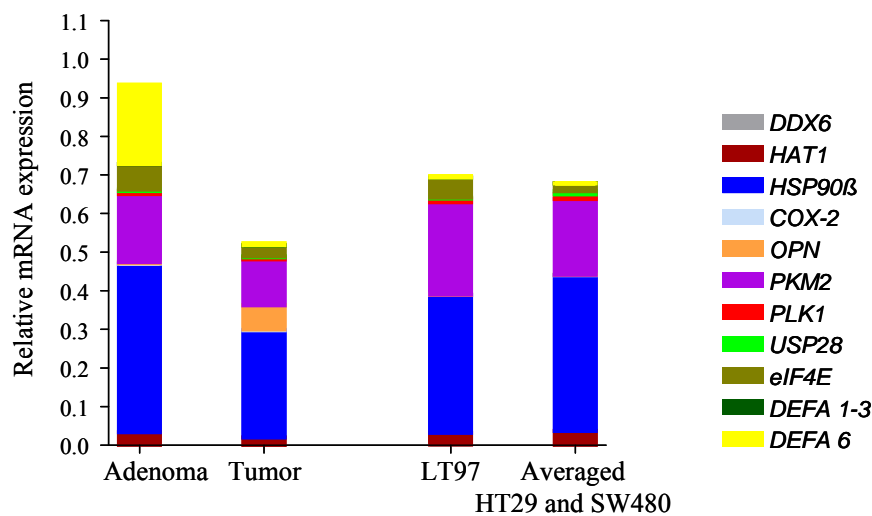
Moreover, it was also observed a deviation in the relative mRNA expression, in the range of 1.2 to 22 fold, between different extractions obtained from one cell line. In this regard, besides the above mentioned explanation, it has also been reported that, with time in culture, Caco 2 undergo spontaneous cell cycle arrests and simultaneous spontaneous differentiation, linked with down-regulation of Wnt/ β -catenin/Tcf signaling (140), which has been suggested to activate the transcription of *DEFA 6* gene. Thus, with the time in culture, Caco 2 can diminish the already low *DEFA 6* expression to levels which are undetectable using this type of assay. A very low, but still detectable expression level was observed in SW620, as in one of the three cell extractions, *DEFA 6* expression was below the detection limit. This result was in agreement with the report from Nam et al., 2005 where the expression of the Paneth specific defensin gene was confirmed in SW480 and SW620 cell lines (93).

Furthermore, since the relative expression of all 12 investigated genes in cell lines as well as in colon tissues was calculated in the same way, using the same normalization control, direct comparison between the gene expression patterns in colon tissues and cell lines was performed. Gene expressions in adenoma and tumor tissue (derived from patients with colorectal cancer) were compared with the respective expressions in LT97 (adenoma), HT29 and SW480 colorectal cancer cell lines. HT29 and SW480 cell lines represent different grades of colorectal cancer and therefore their expression for a defined gene was averaged, in order to be compared, with the evaluated expression in human tumor tissue, which indeed collects cancers with different stages as well. Thus, bar code representation was used to illustrate the combined expression patterns of all investigated genes in a certain type of tissue and/or cell line, see Figure 4-19, page 73.

The analysis revealed big expression differences between the examined adenoma and tumor colon tissue on one hand and the corresponding colon cell lines on the other hand, in respect

to *OPN*, *DEFA 1-3* and *DEFA 6* genes. The relative *OPN* expression in adenoma and tumor tissue was more than 320 fold higher compared with the *OPN* expression in LT97 and the averaged expression in SW480 and HT29 cell lines. As mentioned above, *DEFA 1-3* expression was not detected in LT97, whereas the gene was amplified in adenoma tissue. The presence of *DEFA 1-3* in tissue can be also a consequence of invading neutrophils (92; 99). However, *DEFA 1-3* was found in SW480 and HT29, but it was considerably low (47 fold) compared to the respective expression in patient-derived tumor tissue.

Figure 4-19 Combined relative expression of all investigated genes (see the legend of the figure) in colon tissues obtained from colorectal cancer patients (adenoma and tumor) and cell lines, representing different stages of colon cancer development.



In addition, the expression of *DEFA 6* in adenoma and tumor tissue was much higher (>58 fold) than the respective expression in LT97 and the averaged expression in SW480 and HT29 cell lines.

Furthermore, the rest of the genes (*PKM2*, *eIF4E*, *PLK1*, *DDX6*, *USP28*, *HAT1*, *HSP90β*, *COX-2*) showed comparably the same expression levels in LT97 cell line and adenoma tissue. The observed expression changes were in the range of 2 fold (note that 1 fold indicates no changes in the expression). Higher differences in the expression (of about 3.4 fold) were found when colon cancer tissues were compared with the averaged expression of HT29 and SW480 colon cancer cell lines. This observation could be expected, since the expression data obtained from tumor tissue reflects the expression pattern of pool of cancers with different clinicopathological characteristics (stage, grade, gender of the patients).

Concluding remarks

The comparative analysis of gene expression data derived from human tissues and cell lines revealed that the expression levels of OPN, DEFA 1-3 and DEFA 6 genes in cultured cells differ drastically from those observed in colon tissue samples. Thus, the found discrepancies between the expression patterns have to be further considered if in vitro results are

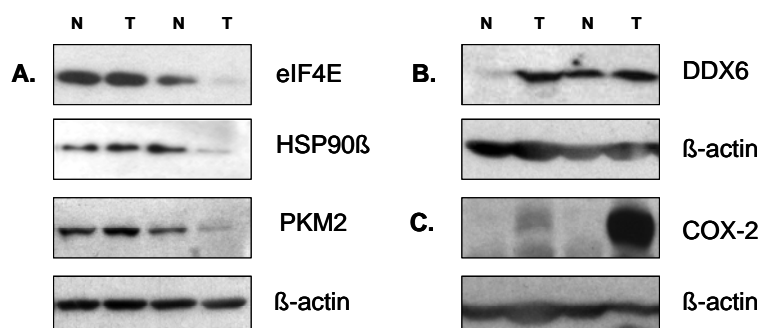
evaluated. In contrast, the rest of the investigated genes showed relatively similar expression levels in cell lines as well as in tissue samples.

4.9. Combined gene – and protein expression analysis of paired normal and tumor tissue, derived from patients with colorectal cancer.

In order to examine whether during cancer development, mRNA abundance of a defined molecule correlates with the respective protein levels, mRNA and protein profiling approaches were applied in parallel.

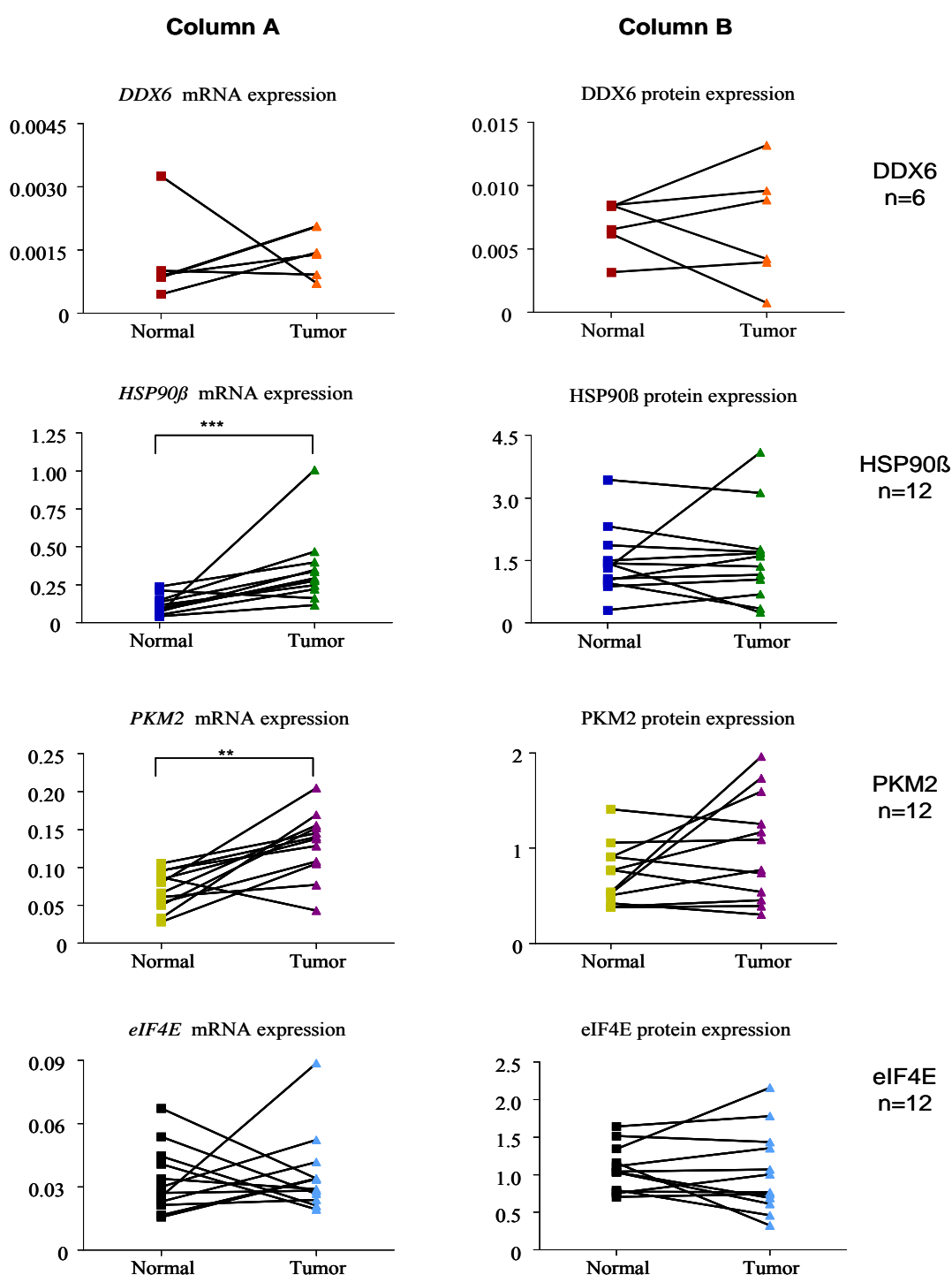
A number of proteins, namely HSP90 β , eIF4E, DDX6, PKM2 and COX-2, were analyzed for their expression profile in paired normal and tumor colon tissues. In total, 12 patients were included in the investigation. However, the computational analysis of DDX6 was performed based only of data derived from 6 out of the 12 tested donors. Representative Western blots are shown on Figure 4-20.

Figure 4-20 Western blot analysis demonstrating the expression of (A.) eIF4E, HSP90 β , PKM2, (B.) DDX6 and (C.) COX-2 in paired normal (N) and tumor (T) colon tissues derived from colorectal cancer patients. Equal protein loading was confirmed with a β -actin antibody.



In order to calculate the expression level of the Western blot-analyzed protein in defined tissue samples, densitometric measurement of its band intensity and the corresponding β -actin (applied as a loading (normalization) control) was performed. Thus, Figure 4-21, page 75, column B showed the computed relative protein expression for each individual patient, in the respective normal and tumor colon tissues. In addition to the relative protein expression, for better comparison, Figure 4-21, column A, also shows the relative gene expression, obtained by qRT-PCR, for paired normal and tumor colon tissues derived from the same patients, who were under Western blot investigation.

Figure 4-21 Relative gene- and protein-expression in paired normal and tumor colon tissue. Column A. Graphs, showing the relative mRNA expression of defined genes across individual patients, Column B. Graphs, showing relative protein expression levels, for each single donor, computed after performance of densitometric analysis of the Western blot. For all experiments, β -actin expression serves as a normalization control. Each symbol represents the relative expression in individual patient as samples which belongs to different tissue groups, but were derived from the same donor are connected to each other. The significant expression differences in tumor with respect to expression in normal tissue are denoted (**, < 0.01 ; ***, < 0.001); n indicates the number of donors included in the analysis.



The significance of the gene- and protein-expression differences between the two analyzed tissue's groups was tested by Wilcoxon matched paired test. Table 4-9 summarizes the respective outcomes.

Table 4-9 Fold changes in the relative expression on gene (column A) and protein (column B) expression levels and the respective outcomes (p-values) from the performed non-parametric Wilcoxon matched pair test. The marked in grey fold changes are significant; n indicates the number of patients included in the computational analysis.

	A. Fold changes in the relative GENE expression in tumor compared to normal tissue	B. Fold changes in the relative PROTEIN expression in tumor compared to normal tissue
DDX6	1.95, p= 0.56, n=6	1.2, p= 1, n=6
HSP90 β	2.97, p= 0.001, *** , n=12	1.00, p= 0.91, n=12
PKM2	1.81, p= 0.0034, ** , n=12	1.11, p= 0.20, n=12
eIF4E	1.08, p= 0.97, n=12	1.01, p= 0.73, n=12
COX-2	26.4, p= 0.0005, *** , n=12	Densitometric measurements not performed

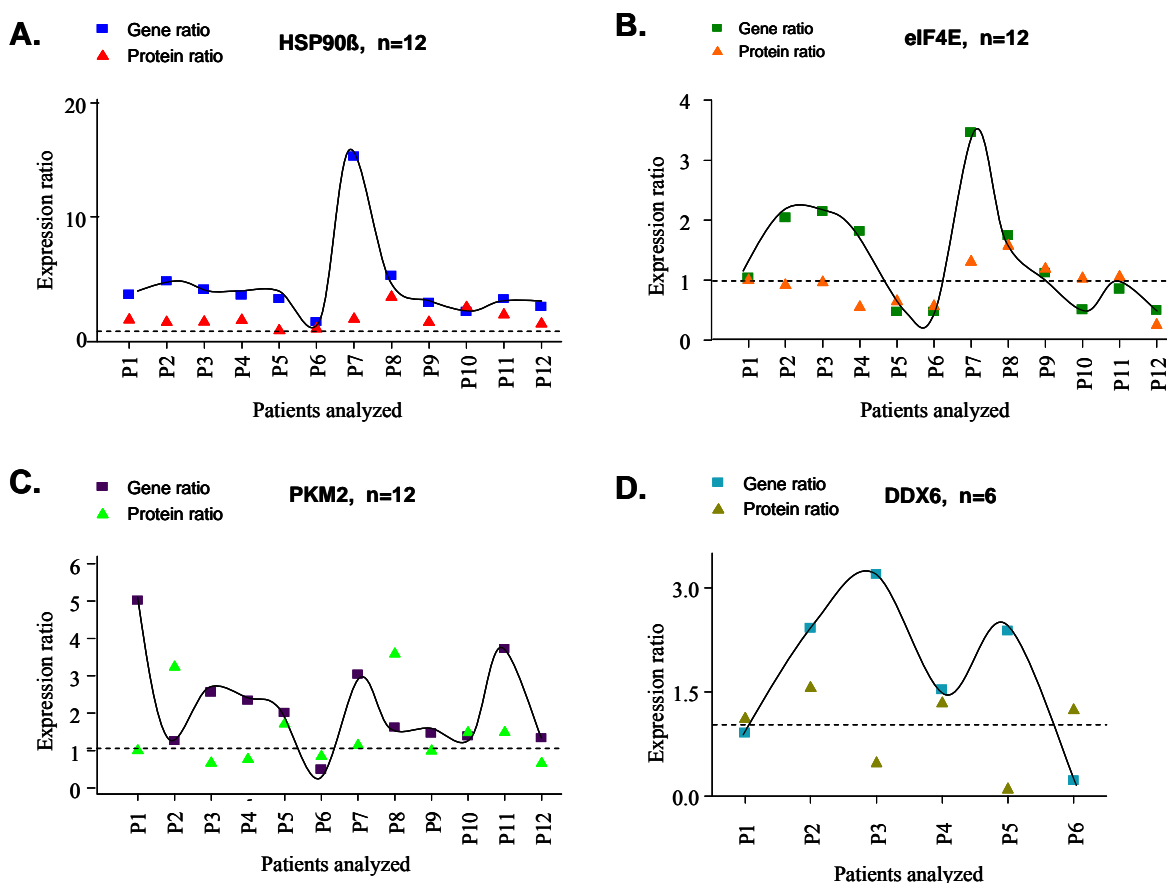
Densitometric measurements of the Western blots and computation of the corresponding expression levels were performed for all proteins included in the comparative analysis, except for the case of COX-2. Only one of the tested 12 donors exhibited extremely high COX-2 expression in tumor tissue (Figure 4-20, C, page 74), the rest of the samples disclosed very weak or not detectable COX-2 signals, especially for samples obtained from normal tissues. Therefore, performance of accurate densitometric measurement was not achievable. As a consequence, the comparison between gene and protein expression levels of COX-2 molecule was not further carried out. However, general elevation of COX-2 protein in tumor was monitored. Similar expression tendency was observed on gene expression level, where *COX-2* mRNA, for the same 12 selected patients, was elevated in tumor (26 fold) compared to related normal tissue. According to the performed statistical test, the increase was significant, (Table 4-9).

Besides the fact that the computed protein expression differences were not significant, the entire subset of investigated proteins did not show any expression alterations in tumor compared to normal tissue. In contrast, *DDX6*, *HSP90 β* and *PKM2* genes revealed slight expression elevation in tumor, as for *HSP90 β* (2.97 fold) and *PKM2* (1.81 fold), the detected expression changes were computed as statistically significant.

Furthermore, the built up gene and protein expression ratios between tumor and normal tissue, for defined molecule of interest, were combined and plotted in a patients-resolved manner (Figure 4-22, page 77). The analysis disclosed that HSP90 β mRNA and protein expression ratios are relatively stable among all patients analyzed (except one patient) (Figure 4-22 A). However, it should be noticed that although stable the HSP90 β gene and proteins expression ratios were in different ranges. While the values for most of the gene expression ratios were above 2, the protein ratios were about 1, which indeed indicated that the protein expressions between normal and tumor tissue did not differ. In addition, the

eIF4E expression ratio revealed considerable inter-individual variability on gene expression level, which was evened down on protein level (Figure 4-22, B). For the rest of the analyzed targets the data was too noisy to be interpreted (Figure 4-22, C and D). Therefore, more experiments and data points are needed in order to confirm and put on a more solid base the already found expression tendencies and/or to be able to make more clear assumptions regarding the molecules which presently did not show any tendency in their gene- and protein- expression patterns.

Figure 4-22 Combined and plotted in a patients-resolved manner are the gene and protein expression ratios between normal and tumor tissue, for (A.) HSP90 β , (B.) eIF4E, (C.) PKM2 and (D.) DDX6 molecules. Each symbol denotes either gene (■) or protein (▲) expression ratio in individual patient (P). For better visualization, the gene expression ratios among the all analyzed patients are connected. Ratio equal to 1 indicates no differences between the expression levels in tumor and normal tissue (broken horizontal line crosses the y-axis at 1); n indicates the number of analyzed patients.



Concluding remarks:

The combined analysis of gene and protein expression data revealed that among all 12 patients tested, the HSP90 β and PKM2 genes showed relatively low, but significant amplification in tumor tissues. On the protein side, however, no changes in the expression were found. Further, when the data were evaluated in a patient-wise manner, the HSP90 β mRNA and protein expression ratios, between normal and tumor tissue, showed to be relatively stable among all analyzed patients. However, the value of these expression ratios differed. In contrast, eIF4E disclosed higher inter-individual variability of the ratios on the mRNA level which was evened down on the protein level. All these observations can be a consequence of either a possible post-transcriptional/translational regulation and/or instability of the protein in the tumor cell or can be due to a smaller sensitivity of the protein assay to detect small expression differences.

5. DISCUSSION

5.1. *β-actin* and *GUS* as housekeeping genes are stably expressed in colon tissues. No need of multiple normalization strategy!

Accurate normalization of the gene expression level is an absolute requirement for reliable results, especially when the biological significance of subtle gene-expression differences is studied. A number of reports have shown already that housekeeping gene expression can vary considerably not only among different types of tissue, but also among different disease states and/or different treatments (118-121). As a consequence, misinterpretation of the outcome can occur (123). Therefore, in order to find the most appropriate internal normalization control, many groups started to evaluate the gene expression variability of classically used housekeeping genes among various types of tissue and disease states (118; 121; 141). Moreover, the application of multi-normalization strategy, where 2 or more housekeeping genes from different functional classes are used, becomes widely used (123). In the present study, apart from the normalization to the amount of total RNA added to each reaction, two widely used housekeeping genes, which are involved in different cellular processes, namely *β-actin* and *β-Glucuronidase (GUS)*, were selected to serve as an internal reference control. Whereas *GUS* takes part in carbohydrate metabolism, the *β-actin* gene corresponds to a molecule comprising part of the cytoskeleton network. This functional independency makes them suitable to be applied as endogenous controls for calibration, since it is likely that they are not co-regulated.

The analysis, carried out in this study showed that apart from their different cellular abundance (Figure 4-1, page 36), *β-actin* and *GUS* possessed minimal variance in their expression patterns across all patients analyzed, independent of the type of colon tissues investigated (Table 4-1, page 38 and Figure 4-2, page 37). This observation is in agreement with previously published data from Rubie et al., where relatively stable expression of *β-actin* and *GUS* in paired tissue derived from colon cancer patients has been demonstrated (118). Similar findings were described by Blanquicett et al. who reported that *GUS* is one of the least variable housekeeping genes among a set of others analyzed endogeneous controls, in paired normal and cancer colon tissues (119). Aerts et al. also demonstrated that apart from 18s rRNA, *β-actin* and *GUS* have the best expression stability across 26 tumor cell lines tested, including colon cancer cell lines as well (142). Further, the performed Pearson correlation analysis displayed that both calibration genes disclosed correlation in their expression patterns among all colon tissue types, except the adenoma colon tissue, which probably can be explained by the small number of samples (n=12) included in the calculation (Figure 4-3, page 39). Moreover, the transcriptional levels of *β-actin* and *GUS* were not significantly influenced by age or gender of the objects, which make them suitable to be used for further normalization of the gene expression data.

When the transcription levels of the gene of interest were normalized to a single or to a combination of the two reference genes, only inconsiderable differences in the computed

outcomes occurred. This observation, together with the above described findings permit the usage of only one single gene for further normalization of the entire set of investigated genes. Thus, in an agreement with many other gene-expression reports (143-145), β -actin was selected to be used as an internal control. This choice was supported by the fact that this HKG also serves as a loading control for protein expression analysis.

Conclusions:

Apart from the lack of 100% reliable internal control, many contradictory observations have been reported regarding the expression stability of one housekeeping gene among different tissue types. These findings emphasize the need to select and validate proper endogeneous control for each tissue under investigation as well as for each experimental condition prior the analysis. Thus, in the present study, the choice for such reliable normalization control was the β -actin molecule, which was used for normalization of the protein expression data as well.

5.2. A burst of alpha defensin 6 expression in adenoma stage of human colon carcinogenesis

One of the main discoveries in this work was the unexpected extremely high increase of *DEFA 6* gene expression in benign adenoma tissues over the normal tissues, obtained from patients with colorectal cancer. With almost 85 fold, *DEFA 6* was shown to be overexpressed in adenoma tissue when compared to the respective normal tissue. In contrast, comparably low was the elevation in its expression in tumor over the normal tissue (5.8 fold), (Table 4-3, page 44).

DEFA 6 as well as the other analyzed neutrophilic molecules (*DEFA 1-3*) belongs to the alpha defensin subfamily. All of these defensins are well known as antimicrobial peptides in the host defense of the gut and none of them was detected in healthy colon, but solely in chronically inflamed colon tissue (90; 94). Under chronic condition, constitutive activation of NF- κ B occurs, which in turn can promote cancer development through enhanced cellular proliferation and inhibition of apoptosis (108). One of the many malignancies known to be initiated by this way is the inflammatory bowel disease (IBD)-associated colon cancer, though with low percentage, but with higher mortality than the sporadic forms (146).

Alpha defensin expression was tested in different types of samples, derived from colorectal cancer patients, such as stool (92), serum, plasma (90; 91) and colon tissue (89; 92; 99). The latter sample type consists of a mix of different cell populations such as epithelial cells, a small percentage of fibroblasts and leukocytes.

Widely spread is the knowledge that neutrophils are the main source of *DEFA 1-3* and this was confirmed in the present study for leukocytes (PBMcs), obtained from healthy individuals, where an expression with several orders of magnitude higher than in all colon tissues and cell lines has been found (Figure 4-8, page 50). Therefore, the assumption that the elevated *DEFA 1-3* expression in tumor tissues may be better explained with the invasion of neutrophil leukocytes rather than with the defensin production in the cancer cells

themselves seems reasonable (92; 99). In contrast to these reports, other research groups showed that although less pronounced than in neutrophils, DEFA 1-3 are also expressed in malignant colon cells (90; 101). However, it is still unclear how the malignant cells themselves influence the elevation of DEFA 1-3 expression in cancer tissues. It was reported, for example that DEFA 1-3 may stimulate bronchial epithelial cells to up-regulate interleukin-8 (IL-8) production, which then can modulate the transcriptional activation of DEFA 1-3. Therefore, it was speculated that DEFA 1-3 increase in tumors may primarily originate from invading immune cells, but initiated by DEFA 1-3 producing cancer cells (91; 147).

Apart from the very high expression in leukocytes (PBMC), the results presented here showed a low, but significant *DEFA 1-3* expression in premalignant, adenoma stage of colon cancer as compared to the corresponding normal tissues (4.01 fold). Increased *DEFA 1-3* transcription was found in tumor tissues as well (6.4 fold). However, our results verified the general observation of *DEFA 1-3* elevation in colon tumors, see Table 4-3, page 44.

In contrast to *DEFA 1-3*, *DEFA 6* mRNA was not detected in PBMCs. This is not surprising, since it is well known that the main producers of DEFA 6 are the Paneth cells, which are also not found in healthy, but only in inflamed and malignant colon and stomach tissues (148; 149). In agreement with the present observations, *DEFA 6* gene was reported to be elevated in colon cancer tissue and some of the cell lines (93). The authors also confirmed this finding on the protein expression level. However, the surprise was the detection of extremely high *DEFA 6* expression in benign adenoma samples. In most patients the effect was reverted in tumor, where *DEFA 6* expression was distinctly lower compared to adenoma, but still remained much higher than in normal tissue. Nonetheless, to our best knowledge this is the first study demonstrating a gene expression explosion of *DEFA 6* in premalignant, benign adenoma.

While the transcriptional activation of *DEFA 1-3* mainly occurs via TLR (NOD2)/NF κ B signaling pathway the Wnt/ β -catenin/Tcf pathway is responsible for the transcriptional activation of *DEFA 6*. These two pathways regulate genes encoding proteins implicated in processes such as cellular proliferation, apoptosis, survival and motility, making them very essential for the development of cancer (1; 6; 25). Therefore, it is expectable that they will be hyperactive during the tumorigenesis. Particularly for colon cancer, it was demonstrated that the early key event in the initiation of the adenoma is the functional modification of components of the Wnt/ β -catenin/Tcf signaling, which also appears to control the intestinal homeostasis. Thus, in more than 80% of the early adenomas derived from sporadic colon cancer cases as well as in all cases of inherited form of colon cancer, familial adenomatous polyposis (FAP), a truncated mutation in the tumor suppressor gene APC was found (140). Some of the remaining cancer cases carried mutations in β -catenin and/or Axin2, other components of the Wnt cascade (25). The immediate consequence of those mutations is hyper activation of the pathway leading not only to promotion of tumor development, but also to transcriptional activation of the *DEFA 6* target gene (6). Therefore, it is reasonable to hypothesize that a *DEFA 6* burst in an early pre-malignant stage of colon cancer

development is a consequence of hyperactive Wnt/ β -catenin/Tcf signaling in intestinal epithelium. However, this assumption does not exclude the possible influence of other signaling cascades on DEFA 6 transcriptional activation. Such cross-talk between the pathways may also be the reason for the reverted *DEFA 6* expression in the latest stage of cancer development, where its transcription can be interfered by other signaling cascades that become modified in a later stage of tumorigenesis. Further examinations, aiming to address these questions, are necessary.

Conclusions:

To our knowledge this is the first study demonstrating a strong burst of the DEFA 6 gene in human adenoma tissue samples, which immediately indicates its potential to be used as a marker for early premalignant stages of colorectal cancer and not solely as a marker for colon cancer detection in general.

5.3. Correlation between the gene expression pattern and clinico-pathologic factors in colon cancer

When working with individuals, it would be reasonable to evaluate the prognostic significance of the gene under investigation. This can be done by testing the correlation between the relative expression of the molecule examined and the clinico-pathological features of the donors such as age, gender, grade and stage of cancer.

5.3.1. The expression of the investigated genes is not associated with the age of the patients

The senescence of the gastrointestinal tract was reported to be associated with processes well known to be implicated in the development and progression of carcinogenesis, namely induced proliferation and apoptosis inhibition. Thus, it was suggested that ageing may predispose the gastrointestinal tract to neoplasia (150).

If age-related transcriptional changes are a consequence of the ageing process, similar alterations might be expected among species, and certainly, organs of the same animal. However, it was demonstrated that many of the age-related transcriptional alterations are not shared between different species and also between different tissues obtained from one species, which underlines the apparent randomness of these changes (151).

Although none of the molecules investigated showed to be significantly dependent on age, many studies already reported such correlation.

COX-2 and ageing

The COX-2 molecule was shown to be constitutively expressed in the colonic wall (39). As enzyme involved in prostaglandin synthesis, COX-2 is implicated in diverse physiological processes such as cell proliferation, survival, angiogenesis as well as inhibition of apoptosis.

Many of those processes are known to change with age. In addition, age-associated alterations in factors well known to induce the *COX-2* gene expression, such as oxidative stress (152) and inflammation (153) can also be regarded to the senescence-induced *COX-2* transcriptional activation. In this line, several reports indicates altered *COX-2* expression with increasing age not only in healthy (154-156), but in cancer tissues as well (128; 157). In contrast, Zhan et al. did not find any correlation between the age of colorectal cancer patients and the *COX-2* expression level (158). This observation is in agreement with the data reported in the present study, where according to the Spearman correlation analysis, *COX-2* was not significantly correlated with age: $p=0.34$ in normal and $p=0.57$ in tumor tissue, (Table 4-4, page 51).

PLK1 and ageing

A progressive decrease in the expression of the cell cycle regulator PLK1 has been reported in actively dividing human fibroblast with age (159), which might contribute to the loss of tissue regenerative capacity. However, in consistence with Macmillan et al. (160), the present study did not show any significant correlation between the expression of PLK1 and the age of the patients when normal colon tissues were examined ($p=0.57$ for normal tissue). It was speculated that the lack of any association between PLK1 expression and the age can be a consequence of the nature of the investigated tissue, namely intestinal epithelial cells which are characterized by high baseline proliferative activity. The colon tissues used in the present study contain of around 75% of epithelial cells (data derived from the PhD work of Julia Sauer, Department of Nutritional Toxicology, Institute for Nutrition, Friedrich-Schiller-University Jena). Therefore, as a crucial regulator for cell division, the PLK1 expression is probably maintained during ageing. Moreover, the Macmillan's study reported an unexpected significant increase of PLK1 expression in tumor tissue with age, which can be caused by the hypothetical loss of common regulatory mechanism with the process of senescence. However, in the present work no correlation with the age has been found ($p=0.44$ for tumor, Table 4-4, page 51).

HSP90 β and ageing

Often, during the ageing process many proteins undergo different modifications, such as oxidation for instance. Those abnormal proteins can serve as stress signals for induction of Heat Shock Proteins (HSPs). The latter are highly inducible molecules which protect the modified proteins from denaturation. While Njemini et al. reported significant increase in HSP90 expression with age in blood cells (monocytes and lymphocytes) from healthy voluntaries, but not in patients with acute infections (161; 162), Boehm et al. as well as other research groups described decreases in HSP90 β mRNA and protein expression with age in numerous cell types. This leads to the hypothesis that diminishing HSP90 levels play a role in the multiple pathways and cellular dysfunctions that characterize senescence and result in the initiation of age-associated diseases (163). In the present work, no association

between HSP90 β expression and age was observed (Spearman analysis, $p=0.43$ in normal and $p=0.33$ in tumor tissues).

eIF4E and ageing

Further, in consistence with Berkel et al. the present study showed that age did not effect the expression of eIF4E along the adeno-carcinoma axis (54). According to the applied Spearman test, no correlation between gene expression and age was found. P-values higher than 0.05 were computed for normal ($p=0.68$) and tumor tissue ($p=0.3$), (Table 4-4, page 51). However, a recent study showed that the depletion of eIF4E, specifically in somatic cells extends the lifespan in *C. elegans* (164). In contrast, Ruggero et al. presented that overexpression of eIF4E induces cellular senescence in mouse embryonic fibroblast (165).

OPN and ageing

Previous studies reported increased OPN expression with the age in neurons (166), mouse macrophages and human kidney proximal tubule epithelial cells (PTE) (167). OPN protein expression was significantly associated with age and grade of cancer in patients with gastric carcinoma (168). In contrast, no significant correlation between age and OPN in patients with non-small-cell lung cancer has been indicated (129). Nonetheless, the analysis illustrated significant correlation with gender of the objects and the grade of cancer.

The above mentioned observations suggest that probably no common correlation between OPN expression and age among the different types of tissues exists. This speculation is reasonable, since OPN has been reported to be expressed in many different tissue types and cells and associated with various cellular processes. However, recently it has been published that ageing of the colonic mucosa is associated with the activation of Phosphatidylinositol 3-kinase /Akt signaling (PI3K/Akt pathway), with resulting transcriptional activation of the downstream effector molecules such as OPN (150; 169). The latter observation leads to the assumption that OPN expression should be positively correlated with age in regard to colonic mucosal-ageing. However, the present study did not find any significant age-related alterations, possibly due to the narrow range of patient's age analyzed (ranged from 58 to 89) and the small sample size.

DEFA 6 and ageing

The present study revealed a slight tendency of positive correlation between the expression of the gene coding for the antimicrobial peptide DEFA 6 and the age of the patients ($p=0.1$ in normal and $p=0.15$ in tumor tissue, see Table 4-4, page 51). In this respect, no other observations in the literature have been reported. However, it would be reasonable to speculate that the age-induced expression of this antimicrobial molecule could be partially a consequence of the paradoxical increase of inflammation (chronic inflammation), a feature commonly observed in ageing (153).

The literature data is in line with the outcomes of the present study, where no association has been found between the expression of DEFA 1-3, DDX6, USP28, HAT1 and PKM2 and the age of the patients, either in normal or in tumor colon tissues.

5.3.2. The expression of the examined genes is not correlated with the grade and the stage of the cancer tissues obtained from individual patients.

In order to develop an individual treatment plan for each analyzed donor, as well as to predict the patient's prognosis, information regarding the grade and the stage of the cancer is necessary.

Tumor grade is a system used to classify cancer cells in terms of how abnormal they look under a microscope and how quickly the tumor is growing and spreading. In contrast, the staging describes the extend or severity of an individual's cancer based on the degree (extend) of the primary tumor and the level of spreading in the body (for more detailed information, refer to: <http://www.cancer.gov/cancertopics/factsheet/detection/tumor-grade>).

Although many expression analyses revealed associations between those clinical parameters and the expression of certain molecules, the present study did not show the existence of any significant correlation, according to the performed Kruskal-Wallis and Spearman correlation analysis.

COX-2, grade and stage of cancer

Fujita et al. suggested that tumors with more COX-2 grow larger and in a more invasive manner (170). Further, it was shown that COX-2 expression was significantly associated with tumor size, depth of invasion, lymph node metastasis, vessels invasion, stage and recurrence (171). In contrast to these reports, Zhan et al. and others as well as the present study did not find such a correlation (158). No significant changes in the expression levels among the different grades and stages of cancer were detected.

PLK1, grade and stage of cancer

The expression of PLK1 as a cell cycle regulator has been reported to be significantly correlated with the histological grade and clinical stage of ovarian cancer (172). Moreover, Weichert et al. have shown that PLK1 expression was positively linked to the stage of colon cancer and nodal status (68). In contrast, Macmillan et al. found a lack of correlation between PLK1 expression and the stage of colorectal cancer, which however makes this molecule a potential independent prognostic marker (160). In consistence with this observation, the present study did not show any relationship between the PLK1 expression levels and the stage of colon cancer.

HSP90 β , grade and stage of cancer

HSP90 and its cochaperones have been shown to facilitate the neoplastic transformations via apoptosis inhibition (81). In this regard, HSP90 chaperon expression has been found to

be associated with a histologic grade of the patients (173) and its overexpression was reported to be connected with a poor patient prognosis. However, the present study did not find any correlations between the expression of *HSP90 β* gene and the grade or the stage of cancer.

eIF4E, grade and stage of cancer

Previous reports showed that eIF4E overexpression can result in malignant transformation in cell lines. Additionally, Berkel et al. demonstrated a highly significant correlation between the malignant potential of the premalignant colon lesions and the level of eIF4E expression. Highest level of it was detected in adenocarcinomas and the lowest levels were found in histologically normal mucosa (8; 54). However, the present study did not find any significant relationship between eIF4E transcriptional level and the various stages of carcinogenic process.

DDX6, grade and stage of cancer

As for the previously described targets, no relation between grade and stage of colon cancer and *DDX6* expression was detected. This observation was in agreement with the data reported from Nakagawa et al. (174).

OPN, grade and stage of cancer

Induced OPN expression was found to be significantly correlated with the stage of cancer, but not with the tumor grade (175; 176). It was speculated that such correlation most likely reflects the effect of OPN on cell migration. OPN can support *in vitro* attachment for a variety of cell types and promotes migration of inflammatory and tumor cells. Moreover, by binding to various cell surface receptors, OPN can induce changes in tumor cell gene expression, including induction of proteolytic enzymes and activation of growth factor kinases, which in turn can lead to increased cell motility and invasion. In addition, *in vivo* experiments have demonstrated that OPN is preferably expressed by tumor cells with invasive and metastatic properties, suggesting that it may be a marker for cancer in a high stage (175). Nonetheless, our computational analysis revealed no correlation between the OPN expression level and the stage of cancer.

PKM2, grade and stage of cancer

In regard to PKM2, strong correlation have been found between the foecal tumor-PKM2 and the stage of colon cancer as well as the grade of the tumor (74; 76; 77; 177). Immunohistological staining of tumor PKM2 in vast numbers of rat and human tumors revealed that increased amounts of the enzyme is a general metabolic alteration during tumorigenesis and correlates with malignancies of the tumor (76). Nonetheless, the present study did not find any significant relation between *PKM2* transcriptional levels and the clinicopathological parameters.

Alpha-defensins, grade and stage of cancer

The expression of DEFA 1-3 has been shown to be elevated in tumor colon tissue in all stages. However, DEFA 1-3 detected in plasma, was increased only in patients with Duke's stage C and D (89). In addition, Holterman et al. observed that the percentage of alpha defensin positive cases increased with the grade of bladder tumors. The invasive cases of that tumor showed higher intensity levels compared to the noninvasive bladder cancer cells (101). In the present study, according to the performed statistical analysis, no significant correlation between alpha defensin transcriptional levels and the clinico-pathological parameters of the cancer was found.

It is highly possible that the size of the available samples is too small in order to find any correlation between expression patterns and the clinicopathological parameters, such as stage and grade.

5.3.3. *COX-2* and *OPN* transcriptional levels are significantly correlated with the gender of the patients under investigation

Several studies have reported that estrogen supplementation reduced the colorectal cancer risk (178; 179). Thus, since the females have a high estrogen background, it was proposed that they are less likely to develop colorectal cancer at all ages than males.

In both genders the normal colon mucosa expresses the two estrogen receptor isoforms, namely ER-alpha and ER-beta. ER-beta is the predominant subtype in colon cells and its decreased levels were associated with colon tumor (180). In this respect, it was shown that re-expression of ER gene, by adding estrogens, can result in growth inhibition of CRC cell lines (181). As a consequence, the variations in the hormone levels between the genders lead to general gender-related expression diversity.

The present study demonstrated a significant increase of *OPN* in colon cancer tissue obtained from male subjects ($p=0.05$). Similar significant increase in the expression of *COX-2* gene in normal tissue, surgically removed from males compared to females, was detected as well ($p=0.02$). The rest of the analyzed molecules did not show any significant alterations with the gender of the objects under investigation (see Table 4-5, page 53).

So far, the available data from the literature concerning estrogen-regulated expression of *OPN* and *COX-2* is controversial. Consistent with the observation from the present study, it was shown that the female sex hormones decrease *OPN* expression in some renal damage states (182) and animal models of vascular injury (183). In contrast, Kayako found that estrogen stimulates the expression of extracellular matrix proteins, including *OPN*, in human alveolar bone-derived cells (184).

In regard to *COX-2*, the results presented here can be easily explained by the already reported observations of *COX-2* down-regulation in estrogen deficient animals when they have been treated with high concentrations of estrogen (185). The same outcome was observed by Zerkovski in breast cancer tissues (186). In contrast, *COX-2* expression was

induced by estrogens in vascular tissues (187). The same inverse effect was shown by Tamura and co-workers in primary human uterine microvascular endothelial cells (HUMEC). They speculate that this effect can be specific since the estrogen was not affecting COX-2 expression in endometrial or endometriotic stromal or epithelial cells (188).

Conclusions:

The above presented data revealed that the gene expression pattern as well as its association with diverse clinicopathological features is specific and depends on the function the molecule has in special cell types and tissues environment. However, reasonable would be to speculate that the small sample size as well as the narrow ageing group available for computational analysis is the main problem for lacking statistically significant observations in such kind of examinations.

5.4. COX-2, OPN, HSP90 β and PKM2 are the most prominent genes for discrimination between normal and tumor colon tissues

Among the entire set of investigated genes, the combination of *OPN*, *COX-2*, *HSP90 β* and *PKM2* genes was identified to perform the best tissue discrimination between non-malignant and malignant colon tissues. This was achieved by newly developed empirical analysis with determined diagnostic power of almost 78% (section 4.4.1, page 56) and widely applied Principal Component Analysis (PCA) where gene expression data collected from 28 non-malignant and 26 malignant colon tissues was assessed (section 4.6, page 61).

According to the published reports, all of the selected molecules revealed to be overexpressed in malignant colon tissue (section 1.4, page 7). However, in the current study not all of them showed a considerable increased expression in tumors. *OPN* was exhibiting the highest expression alteration in tumor tissue, with 13 fold over normal colon tissue, followed by moderately expressed genes such as *DEFA 1-3* (6.4 fold), *DEFA 6* (5.8 fold) and *COX-2* (4 fold). Interestingly, *PKM2* and *HSP90 β* with 1.6 fold and 2.13 fold, respectively, did not show a substantial increase in tumors, Table 4-3, page 44. Though low, their constant expression elevation in tumor cells as compared to their normal counterparts seems to be essential for cancer cell growth and/or survival, which has been speculated for the case of HSP90 by Isaacs et al., (124) and therefore, they might be important for further tissue discrimination.

It should be noted that this empirical diagnostic strategy is in a preliminary stage of development. The number of patients involved in computation of the criteria used for diagnostic performance should be higher. Nevertheless, the outcomes of the assay revealed that this approach successfully discriminates normal from tumor tissue and therefore, in combination with other screening techniques, it can be a powerful tool for the diagnosis of colon cancer. In addition, it was demonstrated that the presentation of the gene expression data for each individual patient as a “bar code” depicts successfully the biological

characteristics of colon cancer and therefore it can be very valuable for a preliminary selection of gene targets with high diagnostic properties.

5.5. Good prediction power of the dbEST data base concerning the gene expression data obtained by qRT-PCR

Another aim of the present work was to test how valuable is the gene expression data derived from NIH's data base dbEST. This database summarizes microarray expression data, so far published in the literature, for normal tissue and their corresponding malignant counterparts (derived from various organs) and makes available such pooled data in a statistically pre-evaluated form (113). In order to evaluate its prediction power, gene expression data derived from colon tissues and generated by qRT-PCR was directly compared with a huge amount of data obtained from NIH's data base dbEST. The comparative analysis revealed that dbEST data base has a good prediction power in respect to paired normal and tumor colon tissues. The expression tendency, for almost all genes, in tumor compared to normal tissue remained the same, independent of the used data source (section 4.7, page 66). Therefore, it is reasonable to assume that the expression prediction for other tissue types, by data mining, might be useful as well. This will allow researchers, with a minimum effort, to check for instance, whether a gene acts as cancer-associated gene only in the experimentally investigated cell lines or tissues, or behaves similarly in a wide variety of tissues and organs. For example, earlier findings based solely on data mining revealed that the expressions of 42, termed "ubiquitous cancer genes", were up-regulated in 36 different types of cancer (113). Two of them, *PKM2* and *HSP90 β* , which have shown overexpression in colon cancer, were also under present investigation. It was demonstrated that dbEST quite accurately predicts the expression of *HSP90 β* . Poorer was the agreement between prediction and experiment for *PKM2* relative gene expression. Nevertheless, according to both dbEST and qRT-PCR sources *PKM2* was increased in cancer compared to normal colon tissue. This places a number of interesting *in silico* investigations, solely based on dbEST data base, on significantly safer ground (189).

One application of the present results for future studies would be to classify the gene expressions in different individuals despite the presence of a strong inter-individual variation. In this respect, a typical example is the *HSP90 β* gene (Figure 4-16, page 67). The normalized expression values for *HSP90 β* were vastly variable among the patients. Apart from the individuals, who revealed *HSP90 β* expression in malignant tissue closely in line with the dbEST prediction, there were patients whose tissue expression deviated significantly from the expression obtained from the data base. This is not surprising if the fact that the dbEST averages over hundreds of patients is taken into account. Therefore, the patients who disclosed expression profile which is in agreement with the dbEST are "typical ones", while those with a high deviation are "outliers". The latter, for example, can be considered to undergo modified treatment of their cancer.

Conclusions:

The present study showed that dbEST gives good predictions of gene expression for comparison with clinical data. This was shown in respect to non-malignant and malignant colon tissues derived from patients with colorectal cancer. Thus, the wealth of gene expression data buried in dbEST is more reliable than one might have expected so far. However, it has still to be verified, for few more tissue types, that the patient's derived gene expression data obtained after performance of qRT-PCR indeed corresponds to the data obtained from dbEST data base or vice versa.

The dbEST database can serve as a good "information source", to perform advanced/preliminary screening for the gene expression pattern of one or a set of genes, involved in one particular or a net of cellular processes over all different types of non-malignant and/or malignant tissues. This useful information could give researches valuable directions for future investigations.

5.6. Messenger RNA (mRNA) expression does only moderately predict protein expression

It is a well known fact that some mRNAs are transcribed, but not translated and therefore the number of mRNA copies does not necessarily reflect the number of functional protein molecules (190), which are mainly responsible for tumor phenotype (191). Moreover, the effects of environmental factors or processes such as nutrition or aging can not be evaluated simply just only by examination of gene expression pattern alone. Therefore, in addition to the valuable gene expression profiling, methods examining the protein profiling in the cells are required too. Such a protein analysis can adequately predict which proteins are expressed, to what quantity, and in what form and thus to complement the gene expression studies. Additionally, the combined gene and protein analytical techniques may define the complex relationship between transcription and translation, and consequently between mRNA and protein. In this regard, however, for humans only few reports are available so far (190; 192). All of them showed discrepancies in the relationship between mRNA and protein expression. Besides positive correlations between mRNA and protein expression levels, moderate and varied correlations were found as well. Therefore, it was speculated that no consistent mRNA-protein correlation exist. The analyses also showed that the correlation strength largely varies among different genes and different biological categories and only a subset of the studied mRNAs, with critical biological functions, had a significant positive correlation with the respective protein expression (190).

Based on patient examinations it was also suggested that mRNA-protein expression correlation might vary in different individuals (190). This was confirmed also by the observation presented in this study where the gene and protein expression ratios, between tumor and normal tissue, were presented in patient-wise manner (Figure 4-22, page 77).

In addition, a precise parallel gene and protein examination can address the question whether the selected set of molecules can efficiently discriminate non – malignant from malignant stage not only on gene, but on protein expression level as well.

The evaluated data, across all analyzed patients, showed that on the gene expression level the HSP90 β and PKM2 molecules exhibited relatively low, but significant up-regulation in cancer compared to the respective normal tissue. On the the protein level, however, no expression alterations were observed either for PKM2 or for HSP90 β molecule. The lack of such correlation between mRNA and protein expression can be caused by many factors. One factor can be the accuracy of the methods used for detection of mRNA and proteins. Another reason which can affect the relative quantities of mRNA and protein for defined molecule are the various complicated biological processes, such as transcriptional and/or post-transcriptional splicing, translational modifications, translational regulation, and protein complex formation. The degradation rates of mRNA and proteins also can affect the estimation of correlation between mRNA and the corresponding protein molecule (190).

Furthermore, the present study showed that across all investigated donors no alterations were observed in regard to the eIF4E molecule on gene and protein expression levels. For such a stable expression correlation between mRNA and protein also the lack of any other known post translational modification of eIF4E, except its phosphorylation at S204, might contribute. This modification enhances the mRNA transport function of eIF4E and its transcription activation in cell culture, but it was not reported to influence the amount of protein in the cell (193).

In respect to the DDX6 molecule only 6 donors were analyzed in parallel for their mRNA and protein expression patterns. When the relative expression in tumor tissue was compared with the respective expression in normal tissue, a moderate correlation between mRNA and protein expression levels was found among all analyzed donors. This observation was in agreement with the report of Akao et al., who showed that gastric mucosa produced a moderate amount of DDX6 protein consistent with the level of DDX6 mRNA (48). They also showed that some of the tissues, such as brain, lung and muscle, produce DDX6 mRNA, but not the encoding protein, which leads to the conclusion that probably post transcriptional regulation of the DDX6 gene expression appears to operate in these tissues. Therefore one can conclude that the regulation of the DDX6 molecule is most likely tissue-specific.

Although in this work a direct comparison between COX-2 mRNA and protein expression levels was not performed (due to technical problems), published data indicate that COX-2 expression is not only regulated on the transcriptional and post-transcriptional level but may also be regulated on the protein level (i.e. by the rate of protein synthesis and/or degradation (by ubiquitin-proteasome pathway)) (194). As a consequence the correlation between mRNA and protein can be affected.

Conclusions:

In conclusion, the mRNA levels often can predict the protein levels, however, this might not to be true for all cases.

6. REFERENCE LIST

1. Sancho E, Batlle E, Clevers H. *Annu Rev Cell Dev Biol* 2004;20:695-723.
2. Bertucci F, Salas S, Eysteries S, et al. *Oncogene* 2004;23:1377-1391.
3. Brennan DJ, O'Brien SL, Fagan A, et al. *Expert Opin Biol Ther* 2005;5:1069-1083.
4. Howe LR, Subbaramaiah K, Chung WJ, et al. *Cancer Res* 1999;59:1572-1577.
5. Rohde F, Rimkus C, Friederichs J, et al. *Int J Cancer* 2007;121:1717-1723.
6. Andreu P, Colnot S, Godard C, et al. *Development* 2005;132:1443-1451.
7. Araki Y, Okamura S, Hussain SP, et al. *Cancer Res* 2003;63:728-734.
8. Rosenwald IB, Chen JJ, Wang S, et al. *Oncogene* 1999;18:2507-2517.
9. Coura RS, Ashton-Prolla P, Prolla JC. *Arq Gastroenterol* 2005;42:99-106.
10. Lynch HT, de la CA. *J Med Genet* 1999;36:801-818.
11. Groden J, Thliveris A, Samowitz W, et al. *Cell* 1991;66:589-600.
12. Kinzler KW, Vogelstein B. *Cell* 1996;87:159-170.
13. Thibodeau SN, Bren G, Schaid D. *Science* 1993;260:816-819.
14. Rampino N, Yamamoto H, Ionov Y, et al. *Science* 1997;275:967-969.
15. Huang J, Papadopoulos N, McKinley AJ, et al. *Proceedings of the National Academy of Sciences of the United States of America* 1996;93:9049-9054.
16. Urbanski SJ, Fogt F. *Int J Surg Pathol* 2000;8:11-16.
17. Fearon ER, Vogelstein B. *Cell* 1990;61:759-767.
18. Boland CR. *Cancer* 1993;71:4180-4186.
19. Rajagopalan H, Nowak MA, Vogelstein B, et al. *Nat Rev Cancer* 2003;3:695-701.
20. Nucci MR, Robinson CR, Longo P, et al. *Hum Pathol* 1997;28:1396-1407.
21. Takahashi M, Mutoh M, Kawamori T, et al. *Carcinogenesis* 2000;21:1319-1327.
22. Iacopetta B. *Human Mutation* 2003;21:271-276.
23. Lynch HT, de la CA. *N Engl J Med* 2003;348:919-932.
24. Baylin SB, Ohm JE. *Nat Rev Cancer* 2006;6:107-116.
25. Gregorieff A, Clevers H. *Genes Dev* 2005;19:877-890.
26. Behrens J, Jerchow BA, Wurtele M, et al. *Science* 1998;280:596-599.
27. Ikeda S, Kishida S, Yamamoto H, et al. *Embo Journal* 1998;17:1371-1384.

28. Narayan S, Roy D. *Mol Cancer* 2003;2:41.
29. Wai PY, Kuo PC. *J Surg Res* 2004;121:228-241.
30. Rollo EE, Hempson SJ, Bansal A, et al. *Journal of Virology* 2005;79:3509-3516.
31. Rangaswami H, Bulbule A, Kundu GC. *Trends in Cell Biology* 2006;16:79-87.
32. Junaid A, Moon MC, Harding GE, et al. *Am J Physiol Cell Physiol* 2007;292:C919-C926.
33. Jain S, Chakraborty G, Kundu GC. *Cancer Res* 2006;66:6638-6648.
34. Irby RB, McCarthy SM, Yeatman TJ. *Clin Exp Metastasis* 2004;21:515-523.
35. Agrawal D, Chen T, Irby R, et al. *Comptes Rendus Biologies* 2003;326:1041-1043.
36. Yeatman TJ, Chambers AF. *Clinical & Experimental Metastasis* 2003;20:85-90.
37. Sano H, Kawahito Y, Wilder RL, et al. *Cancer Research* 1995;55:3785-3789.
38. Chapple KS, Cartwright EJ, Hawcroft G, et al. *American Journal of Pathology* 2000;156:545-553.
39. Bernardini N, Colucci R, Mattii L, et al. *Neurogastroenterol Motil* 2006;18:654-662.
40. Eberhart CE, Coffey RJ, Radhika A, et al. *Gastroenterology* 1994;107:1183-1188.
41. Kutchera W, Jones DA, Matsunami N, et al. *Proceedings of the National Academy of Sciences of the United States of America* 1996;93:4816-4820.
42. Kargman SL, O'Neill GP, Vickers PJ, et al. *Cancer Res* 1995;55:2556-2559.
43. Wendum D, Masliah J, Trugnan G, et al. *Virchows Archiv* 2004;445:327-333.
44. Tsujii M, Dubois RN. *Cell* 1995;83:493-501.
45. Tsujii M, Kawano S, Tsuji S, et al. *Cell* 1998;93:705-716.
46. Tsujii M, Kawano S, Dubois RN. *Proc Natl Acad Sci U S A* 1997;94:3336-3340.
47. Linder P, Lasko PF, Ashburner M, et al. *Nature* 1989;337:121-122.
48. Akao Y, Marukawa O, Morikawa H, et al. *Cancer Res* 1995;55:3444-3449.
49. Hashimoto K, Nakagawa Y, Morikawa H, et al. *Carcinogenesis* 2001;22:1965-1970.
50. Akao Y, Yoshida H, Matsumoto K, et al. *Genes Cells* 2003;8:671-676.
51. He TC, Sparks AB, Rago C, et al. *Science* 1998;281:1509-1512.
52. Rosenwald IB, Rhoads DB, Callanan LD, et al. *Proc Natl Acad Sci U S A* 1993;90:6175-6178.
53. Graff JR, Zimmer SG. *Clinical & Experimental Metastasis* 2003;20:265-273.
54. Berkel HJ, Turbat-Herrera EA, Shi RH, et al. *Cancer Epidemiology Biomarkers & Prevention* 2001;10:663-666.
55. de Benedetti A, Harris AL. *Int J Biochem Cell Biol* 1999;31:59-72.

56. McGraw S, Robert C, Massicotte L, et al. *Biology of Reproduction* 2003;68:383-389.
57. Vettese-Dadey M, Grant PA, Hebbes TR, et al. *EMBO J* 1996;15:2508-2518.
58. Verreault A, Kaufman PD, Kobayashi R, et al. *Curr Biol* 1998;8:96-108.
59. Kelly TJ, Qin S, Gottschling DE, et al. *Mol Cell Biol* 2000;20:7051-7058.
60. Seiden-Long IM, Brown KR, Shih W, et al. *Oncogene* 2006;25:91-102.
61. Pogribny IP, Tryndyak VP, Muskhelishvili L, et al. *J Nutr* 2007;137:216S-222S.
62. Gibbons RJ. *Human Molecular Genetics* 2005;14:R85-R92.
63. Arango D, Mariadason JM, Wilson AJ, et al. *Br J Cancer* 2003;89:1757-1765.
64. Popov N, Wanzel M, Madiredjo M, et al. *Nature Cell Biology* 2007;9:765-U71.
65. Toyoshima-Morimoto F, Taniguchi E, Nishida E. *Embo Reports* 2002;3:341-348.
66. Yuan JP, Eckerdt F, Bereiter-Hahn J, et al. *Oncogene* 2002;21:8282-8292.
67. Hansen DV, Hsu JY, Kaiser BK, et al. *Oncogene* 2002;21:6209-6221.
68. Weichert W, Kristiansen G, Schmidt M, et al. *World J Gastroenterol* 2005;11:5644-5650.
69. Smith MR, Wilson ML, Hamanaka R, et al. *Biochemical and Biophysical Research Communications* 1997;234:397-405.
70. Strebhardt K, Ullrich A. *Nat Rev Cancer* 2006;6:321-330.
71. Takahashi T, Sano B, Nagata T, et al. *Cancer Sci* 2003;94:148-152.
72. Ando K, Ozaki T, Yamamoto H, et al. *J Biol Chem* 2004;279:25549-25561.
73. Zhang B, Chen JY, Chen DD, et al. *World J Gastroenterol* 2004;10:1643-1646.
74. Mazurek S, Boschek CB, Hugo F, et al. *Seminars in Cancer Biology* 2005;15:300-308.
75. Mazurek S, Grimm H, Boschek CB, et al. *Br J Nutr* 2002;87 Suppl 1:S23-S29.
76. Hardt PD, Mazurek S, Toepler M, et al. *Br J Cancer* 2004;91:980-984.
77. Tonus C, Neupert G, Sellinger M. *World J Gastroenterol* 2006;12:7007-7011.
78. Hacker HJ, Steinberg P, Bannasch P. *Carcinogenesis* 1998;19:99-107.
79. Schneider J, Neu K, Grimm H, et al. *Anticancer Research* 2002;22:311-318.
80. Takayama S, Reed JC, Homma S. *Oncogene* 2003;22:9041-9047.
81. Bagatell R, Whitesell L. *Molecular Cancer Therapeutics* 2004;3:1021-1030.
82. Isaacs JS, Jung YJ, Mimnaugh EG, et al. *J Biol Chem* 2002;277:29936-29944.
83. Whitesell L, Lindquist SL. *Nat Rev Cancer* 2005;5:761-772.
84. Schmitt E, Gehrmann M, Brunet M, et al. *J Leukoc Biol* 2007;81:15-27.

85. Cen H, Zheng S, Fang YM, et al. *World J Gastroenterol* 2004;10:3122-3126.
86. Kuwano K, Tanaka N, Shimizu T, et al. *Current Microbiology* 2006;52:435-438.
87. Cunliffe RN, Kamal M, Rose FRAJ, et al. *Journal of Clinical Pathology* 2002;55:298-304.
88. Menendez A, Finlay BB. *Current Opinion in Immunology* 2007;19:385-391.
89. Albrethsen J, Moller CH, Olsen J, et al. *European Journal of Cancer* 2006;42:3057-3064.
90. Melle C, Ernst G, Schimmel B, et al. *Gastroenterology* 2005;129:66-73.
91. Albrethsen J, Bogebo R, Gammeltoft S, et al. *Bmc Cancer* 2005;5.
92. Zou H, Harrington JJ, Sugumar A, et al. *Clinical Gastroenterology and Hepatology* 2007;5:865-868.
93. Nam MJ, Kee MK, Kuick R, et al. *Journal of Biological Chemistry* 2005;280:8260-8265.
94. Cunliffe RN, Mahida YR. *J Leukoc Biol* 2004;75:49-58.
95. Hill CP, Yee J, Selsted ME, et al. *Science* 1991;251:1481-1485.
96. Linzmeier RM, Ganz T. *Genomics* 2005;86:423-430.
97. Muller CA, Markovic-Lipkovski J, Klatt T, et al. *American Journal of Pathology* 2002;160:1311-1324.
98. Halder TM, Bluggel M, Heinzl S, et al. *Blood* 2000;95:2890-2896.
99. Zhang K, Lu QH, Zhang Q, et al. *Biochemical and Biophysical Research Communications* 2004;323:437-444.
100. Xu N, Wang YS, Pan WB, et al. *Mol Cancer Ther* 2008;7:1588-1597.
101. Holterman DA, Diaz JI, Blackmore PF, et al. *Urol Oncol* 2006;24:97-108.
102. Richter M, Jurek D, Wrba F, et al. *Eur J Cancer* 2002;38:1937-1945.
103. Laferte S, Loh LC. *Biochem J* 1992;283 (Pt 1):193-201.
104. Ilyas M, Tomlinson IPM, Rowan A, et al. *Proceedings of the National Academy of Sciences of the United States of America* 1997;94:10330-10334.
105. Fogh J, Fogh JM, Orfeo T. *J Natl Cancer Inst* 1977;59:221-226.
106. Rajput A, Dominguez SM, I, Rose R, et al. *J Surg Res* 2008;147:276-281.
107. Ebert MN, Klinder A, Peters WHM, et al. *Carcinogenesis* 2003;24:1637-1644.
108. Daher KA, Lehrer RI, Ganz T, et al. *Proceedings of the National Academy of Sciences of the United States of America* 1988;85:7327-7331.
109. Boguski MS, Lowe TMJ, Tolstoshev CM. *Nature Genetics* 1993;4:332-333.
110. Lal A, Lash AE, Altschul SF, et al. *Cancer Research* 1999;59:5403-5407.
111. Zhang L, Ma XL, Zhang Q, et al. *Gene* 2001;267:193-200.

112. Altenberg B, Greulich KO. *Genomics* 2004;84:1014-1020.
113. Altenberg B, Gemuend C, Greulich KO. *Proteomics* 2006;6:67-71.
114. Altenberg B, Rapp A, Schmitt E, et al. *Genomics* 2007;90:661-673.
115. Davis JC. *Statistics and Data Analysis in Geology*. New York: John Wiley & Sons, 1986.
116. Harper DATed. *Numerical Palaeobiology*. New York.: John Wiley & Sons, 1999.
117. Pfaffl MW, Horgan GW, Dempfle L. *Nucleic Acids Research* 2002;30.
118. Rubie C, Kempf K, Hans J, et al. *Molecular and Cellular Probes* 2005;19:101-109.
119. Blanquicett C, Johnson MR, Heslin M, et al. *Analytical Biochemistry* 2002;303:209-214.
120. Andersen CL, Jensen JL, Orntoft TF. *Cancer Research* 2004;64:5245-5250.
121. de Kok JB, Roelofs RW, Giesendorf BA, et al. *Laboratory Investigation* 2005;85:154-159.
122. Erickson HS, Albert PS, Gillespie JW, et al. *Laboratory Investigation* 2007;87:951-962.
123. Vandesompele J, De Preter K, Pattyn F, et al. *Genome Biol* 2002;3:RESEARCH0034.
124. Isaacs JS, Xu WP, Neckers L. *Cancer Cell* 2003;3:213-217.
125. Deng G, Lu Y, Zlotnikov G, et al. *Science* 1996;274:2057-2059.
126. Aldred PMR, Hollox EJ, Armour JAL. *Human Molecular Genetics* 2005;14:2045-2052.
127. Ballana E, Gonzalez JR, Bosch N, et al. *Bmc Genomics* 2007;8.
128. Badawi AF, Liu Y, Eldeen MB, et al. *Carcinogenesis* 2004;25:1681-1688.
129. Boldrini L, Donati V, Dell'Omodarme M, et al. *Br J Cancer* 2005;93:453-457.
130. Fidler IJ. *Cancer Research* 1978;38:2651-2660.
131. Komori T, Takemasa I, Yamasaki M, et al. *International Journal of Oncology* 2008;32:367-375.
132. Finak G, Sadekova S, Pepin F, et al. *Breast Cancer Research* 2006;8.
133. Koshiji M, Yonekura Y, Saito T, et al. *Annals of Surgical Oncology* 2002;9:580-586.
134. Denhardt DT, Mistretta D, Chambers AF, et al. *Clinical & Experimental Metastasis* 2003;20:77-84.
135. Weber GF. *Biochimica et Biophysica Acta-Reviews on Cancer* 2001;1552:61-85.
136. Marnett LJ, Dubois RN. *Annual Review of Pharmacology and Toxicology* 2002;42:55-+.
137. Lenaerts K, Bouwman FG, Lamers WH, et al. *BMC Genomics* 2007;8:91.
138. Ross DT, Scherf U, Eisen MB, et al. *Nat Genet* 2000;24:227-235.
139. Agerberth B, Charo J, Werr J, et al. *Blood* 2000;96:3086-3093.
140. Mariadason JM, Bordonaro M, Aslam F, et al. *Cancer Res* 2001;61:3465-3471.

141. Radonic A, Thulke S, Mackay IM, et al. *Biochemical and Biophysical Research Communications* 2004;313:856-862.
142. Aerts JL, Gonzales MI, Topalian SL. *Biotechniques* 2004;36:84-1.
143. Tomonaga T, Matsushita K, Ishibashi M, et al. *Cancer Research* 2005;65:4683-4689.
144. Johnston PG, Lenz HJ, Leichman CG, et al. *Cancer Research* 1995;55:1407-1412.
145. Kato H, Semba S, Miskad UA, et al. *Clinical Cancer Research* 2004;10:7318-7328.
146. Baisse B, Fontollet C, Bian YS, et al. *J Clin Pathol* 2004;57:885-887.
147. vanWetering S, MannesseLazeroms SPG, VanSterkenburg MAJA, et al. *American Journal of Physiology-Lung Cellular and Molecular Physiology* 1997;16:L888-L896.
148. Cunliffe RN. *Mol Immunol* 2003;40:463-467.
149. Gassler N, Schnolzer M, Rohr C, et al. *Lab Invest* 2002;82:1647-1659.
150. Majumdar AP, Du J. *Am J Physiol Gastrointest Liver Physiol* 2006;290:G49-G55.
151. Oberdoerffer P, Sinclair DA. *Nat Rev Mol Cell Biol* 2007;8:692-702.
152. Tsurudome Y, Hirano T, Hirata K, et al. *J Gerontol A Biol Sci Med Sci* 2001;56:B483-B485.
153. Gupta S, Agrawal A, Agrawal S, et al. *Immun Ageing* 2006;3:5.
154. Kang KB, Van Der ZA, Iannazzo L, et al. *Transl Res* 2006;148:289-294.
155. Kim JW, Baek BS, Kim YK, et al. *J Gerontol A Biol Sci Med Sci* 2001;56:B350-B355.
156. Wang X, Shen CL, Dyson MT, et al. *Endocrinology* 2005;146:4202-4208.
157. Siironen P, Ristimäki A, Nordling S, et al. *Histopathology* 2004;44:490-497.
158. Zhan J, Liu JP, Zhu ZH, et al. *Chin Med J (Engl)* 2004;117:1151-1154.
159. Ly DH, Lockhart DJ, Lerner RA, et al. *Science* 2000;287:2486-2492.
160. Macmillan JC, Hudson JW, Bull S, et al. *Ann Surg Oncol* 2001;8:729-740.
161. Njemini R, Bautmans I, Lambert M, et al. *Mechanisms of Ageing and Development* 2007;128:450-454.
162. Njemini R, Lambert M, Demanet C, et al. *Biogerontology* 2007;8:353-364.
163. Boehm AK, Seth M, Mayr KG, et al. *Arthritis and Rheumatism* 2007;56:2335-2343.
164. Syntichaki P, Troulinaki K, Tavernarakis N. *Nature* 2007;445:922-926.
165. Ruggero D, Montanaro L, Ma L, et al. *Nat Med* 2004;10:484-486.
166. Wung JK, Perry G, Kowalski A, et al. *Curr Alzheimer Res* 2007;4:67-72.
167. Hwang SM, Lopez CA, Heck DE, et al. *J Biol Chem* 1994;269:711-715.
168. Higashiyama M, Ito T, Tanaka E, et al. *Ann Surg Oncol* 2007;14:3419-3427.

169. Packer L, Pavey S, Parker A, et al. *Carcinogenesis* 2006;27:1778-1786.
170. Fujita T, Matsui M, Takaku K, et al. *Cancer Res* 1998;58:4823-4826.
171. Soumaoro LT, Uetake H, Higuchi T, et al. *Clin Cancer Res* 2004;10:8465-8471.
172. Takai N, Miyazaki T, Fujisawa K, et al. *Cancer Lett* 2001;164:41-49.
173. Yano M, Naito Z, Tanaka S, et al. *Jpn J Cancer Res* 1996;87:908-915.
174. Nakagawa Y, Morikawa H, Hirata I, et al. *British Journal of Cancer* 1999;80:914-917.
175. Coppola D, Szabo M, Boulware D, et al. *Clin Cancer Res* 2004;10:184-190.
176. Agrawal D, Chen TG, Irby R, et al. *Journal of the National Cancer Institute* 2002;94:513-521.
177. Weinberger R, Appel B, Stein A, et al. *Eur J Cancer Care (Engl)* 2007;16:333-337.
178. Martinez ME, Grodstein F, Giovannucci E, et al. *Cancer Epidemiol Biomarkers Prev* 1997;6:1-5.
179. Calle EE, Miracle-McMahill HL, Thun MJ, et al. *J Natl Cancer Inst* 1995;87:517-523.
180. Campbell-Thompson M, Lynch IJ, Bhardwaj B. *Cancer Res* 2001;61:632-640.
181. Issa JP, Ottaviano YL, Celano P, et al. *Nat Genet* 1994;7:536-540.
182. Xie Y, Sakatsume M, Nishi S, et al. *Kidney Int* 2001;60:1645-1657.
183. Li G, Chen YF, Kelpke SS, et al. *Circulation* 2000;101:2949-2955.
184. Matsuta Kayoko. *Nihon University Dental Journal* 2001;VOL.75:11-19.
185. Shin JI KTPO. *J Kor Ass Cancer Prevention* 2003;8:73-80.
186. Zerkowski MP, Camp RL, Burtress BA, et al. *Cancer Invest* 2007;25:19-26.
187. Akarasereenont P, Techatraisak K, Thaworn A, et al. *Inflamm Res* 2000;49:460-465.
188. Tamura M, Deb S, Sebastian S, et al. *Fertil Steril* 2004;81:1351-1356.
189. Radeva M, Hofmann T, Altenberg B, et al. *Curr Pharm Biotechnol* 2008;9:510-515.
190. Guo Y, Xiao P, Lei S, et al. *Acta Biochim Biophys Sin (Shanghai)* 2008;40:426-436.
191. Ogata S, Ho I, Chen A, et al. *Cancer Res* 1995;55:1869-1874.
192. Chen G, Gharib TG, Huang CC, et al. *Mol Cell Proteomics* 2002;1:304-313.
193. Topisirovic I, Ruiz-Gutierrez M, Borden KL. *Cancer Res* 2004;64:8639-8642.
194. Shao JY, Sheng HM, Inoue H, et al. *Journal of Biological Chemistry* 2000;275:33951-33956.

Curriculum Vitae

PERSONAL DATA

Name: Maria Yosifova Radeva

Date and place of birth: 05.02.1977, Sofia, Bulgaria

Nationality: Bulgarian

Marital status: Single

EDUCATION

2004 - 2008 – PhD student in Biology, Friedrich-Schiller-Universität Jena; Adviser: Prof. Karl-Otto Greulich

1995 - 2000 – M. Sc. in Molecular Biology, University of Sofia „St. Kliment Ohridsky”, Sofia, Bulgaria, Faculty of Biology, Laboratory of virology

1991 – 1995 – 22nd secondary school “Georgi.S.Rakovski”; specialty in biology and chemistry, Sofia

EMPLOYMENT

February 2004 – December 2008 – PhD thesis project; Friedrich-Schiller-Universität, Jena and Leibniz Institute for Age Research – Fritz Lipmann Institute (FLI), Germany: Direct supervisor: Prof. Karl-Otto Greulich

May 2002 – December 2003 – Research Assistant; Laboratory of Bioelectrochemistry, Jena, Germany; Direct supervisor: Prof. Hermann Berg

May 2000 – December 2002 – Technical Assistant; Bulgarian Academy of Science, Sofia, Bulgaria; Institute of Biophysics, Department of Excitable Structures; Direct supervisor: Nicolina Radicheva M.D., Ph.D.

PUBLICATIONS

M. Radeva, T.Hofmann, B.Altenberg, H.Mothes, K.K.Richter, B.Pool-Zobel, and K.O.Greulich. The database dbEST correctly predicts gene expression in colon cancer patients. Curr. Pharm. Biotechnol. 9: (2008). 510-515.

M. Lambrea, B. Glück, **M. Radeva**, H. Berg. Electroporation of Cell Membranes Supporting Penetration of Photodynamic Active Macromolecular Chromophore Dextran; Bioelectrochemistry 62 (2004) 95-98

M. Radeva, M. Lambrera, P. Angelova, N. Traitcheva, Hermann Berg. Synergisms of Electric or Electromagnetic Fields and Photodynamic Effects Induce Apoptosis and Necrosis of Cancer Cells; *Electrochemistry*, 10 (2004) 260-270

M. Radeva, A. Berg, H. Berg. Induction of Apoptosis and Necrosis in Cancer Cells by Electric Fields, Electromagnetic Fields, and Photodynamically Active Quinoids; *Electromagnetic Biology and Medicine*, 2004, 23 (3), 185-200

M. Radeva, H. Berg. Differences in lethality between cancer cells and human lymphocytes caused by LF-electromagnetic fields; *Bioelectromagnetics*, 2004, 25 (7), 503-507

H. Berg, **M. Radeva**. Apoptose-Induction und Nekrose bei humanen Krebszellen durch elektromagnetische Felder im Vergleich mit gesunden Lymphozyten.; *Deutsche Zeitschrift für Onkologie* 35 (2003) 75-77

N. Traitcheva, P. Angelova, **M. Radeva**, H. Berg. ELF Fields and Photooxidation Yielding Lethal Effects on Cancer Cells; *Bioelectromagnetics* 24 (2003);148 -150

S. Shishkov, **M. Radeva**, R. Vassileva, T. Popova, Y. Michailov. Comparative effect of OXADIN® against replication of herpes viruses; *Comptes rendus de l'Acad. Bulg.Sci.*, 54 (2001), 73-78

R. Ivanov, P. Naumova, **M. Radeva**, T. Vukova, N. Radicheva. Effect of millimetre wave electromagnetic field on acetylcholinesterase activity of frog skeletal muscle; *Comptes rendus de l'Acad. Bulg.Sci.*, Tome 54 (2001), No 11

MEETING ABSTRACTS:

Maria Radeva, Dietmar Knoll, Thomas Hofmann, Corinna Schwung, Julia Sauer, Wilbert H.M. Peters, Beatrice Pool Zobel and Karl Otto Greulich. Poster presentation: First steps towards multiprotein analysis of colon cancer progression using a Nexterion Slide H. Statusseminar Chiptechnologien, Microarray-Anwendungen in Grundlagenforschung und Routinebetrieb. 1-2 February 2007, DECHEMA

Maria Radeva, Julia Sauer, Thomas Hofmann, Corinna Schwung, Dietmar Knoll, Beatrice Pool-Zobel and Karl Otto Greulich. Poster presentation: Combined analysis of gene- and protein expression data to characterize colon carcinoma, FLI Retreat – 04-06.06. 2007; Bad Blankenburg

M. Radeva, T. Hofmann, J. Sauer, B. Altenberg, B. Pool-Zobel and K.O.Greulich. Poster presentation: Combined analysis of gene- and protein expression data to characterize colon carcinoma, Proceedings of the fall meeting, German Society for Biochemistry and Molecular Biology (GBM), Hamburg, Germany, September 16-19, 2007; September 16-19, 2007, Universität Hamburg

Peter Schellenberg, Hans-Martin Striebel, Paulius Grigaravicius, **Maria Radeva** and Karl Otto Greulich, Antibody-chip analysis using intrinsic fluorescence lifetimes; Single-Molecule Spectroscopy meets Chip-Technology; 18-20 October, 2006

M. Radeva, Poster presentation: First steps towards multiprotein analysis of colon cancer progression. FLI Retreat – 16-18.06.2006; Bad Blankenburg

M. Radeva, R. Vassileva, Laboratory of Virology, Faculty of Biology, University of Sofia, 18-20 May, 2000. Comparative effect of OXADIN against the replication of herpes viruses in vitro; National Student Scientific Session 2000

R. Vassileva, **M. Radeva**, Laboratory of Virology, Faculty of Biology, University of Sofia, 18-20 May, 2000. Inhibitory effect of natural compounds against the replication of HSV-1 and HSV-2 in vitro; National Student Scientific Session 2000

Selbstständigkeitserklärung

Ich erkläre hiermit an Eides Statt, dass ich die vorliegende Arbeit selbstständig angefertigt, keine anderen als die angegebenen Quellen benutzt, die wörtlich oder dem Inhalt nach aus fremden Arbeiten entnommenen Stellen, bildlichen Darstellungen und dergleichen als solche genau kenntlich gemacht und keine unerlaubte fremde Hilfe in Anspruch genommen habe.

Die Promotionsordnung der Biologisch-Pharmazeutischen Fakultät der Friedrich Schiller-Universität ist mir bekannt.

Ich versichere außerdem, dass ich mich mit der vorliegenden Arbeit an keiner anderen Hochschule um den akademischen Grad *Doctor rerum naturalium* beworben habe und dass ich weder früher noch gegenwärtig die Eröffnung eines Verfahrens zum Erwerb des o.g. akademischen Grades an einer anderen Hochschule beantragt habe.

Maria Radeva

Jena, 2009